NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

THE IMPORTANCE OF PHASE IN WORD RECOGNITION

by

Jeffrey T. Pfeiffer
September 1983

Thesis Advisor: S. Jauregui

Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (*and Subtitle*)
The Importance of Phase in Word Recognition

5. TYPE OF REPORT & PERIOD COVERED
Master's Thesis
September 1983

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(*s*)
Jeffrey T. Pfeiffer

8. CONTRACT OR GRANT NUMBER(*s*)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93943

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93943

12. REPORT DATE
September 1983

13. NUMBER OF PAGES
108 pages

14. MONITORING AGENCY NAME & ADDRESS (*if different from Controlling Office*)

15. SECURITY CLASS. (*of this report*)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)
Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)
Phase Representation, Cepstral Processing, Homomorphic Processing, Hilbert Transform, Speech Recognition, Analytic Signal, Short-term Phase, Cepstrum

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

The use of phase-only representations of speech for isolated word recognition is explored. Until recently the ear was thought to be short-term phase insensitive. However, short-term phase-only reconstructed speech has been shown to retain much of the intelligibility of the original signal. Using cepstral and analytic-signal processing techniques, a

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

20. (Continued)

system for isolated word recognition is developed. The results of tests for both the speaker-dependent and speaker-independent case indicate that phase may be an important feature to consider in the development of word recognition systems.

The Importance of Phase in Word Recognition

by

Jeffrey T. Pfeiffer
Lieutenant, United States Navy
B.S., The Pennsylvania State University, 1974

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
September 1983

ABSTRACT

The use of phase-only representations of speech for isolated word recognition is explored. Until recently the ear was thought to be short-term phase insensitive. However, short-term phase-only reconstructed speech has been shown to retain much of the intelligibility of the original signal. Using cepstral and analytic-signal processing techniques, a system for isolated word recognition is developed. The results of tests for both the speaker-dependent and speaker-independent case indicate that phase may be an important feature to consider in the development of word recognition systems.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I wish to express my gratitude and appreciation to my friend, LT Jay H. Benson, for the many hours of enlightening conversation.

I would also like to thank the staff of the W.R. Church Computer Center, especially Messrs. Hillary, LaMont and Mar for the many hours of help they have given me.

A special thanks to Professors C. W. Therrien, D. E. Kirk, and R. D. Strum for introducing me to the subject.

A special thanks to my advisor, Professor S. Jauregui, who made this a most worthwhile educational experience.

Most importantly I wish to thank my loving wife, Carol, who makes every day worth living.

I. INTRODUCTION

As the complexity of man's machines increases, so does the need for simple, efficient man-machine interfaces. Automatic speech recognition plays a major role in this man-machine communication because of the superiority of speech over other modes of human communication. Speech is the most familiar and most convenient way for humans to communicate. Voice input leaves the hands and eyes of the operator free to perform other tasks and allows speaker mobility.

Word recognition is one facet of the research conducted in the area of speech processing. Speech processing can be divided into three major categories. The speech analysis area includes word recognition, speaker identification, and speaker verification. The second category is speech synthesis. An example of synthesis is a data-retrieval system, where the computer responds verbally when its data base is interrogated. Another example is when a child receives a verbal response from his toy informing him he has correctly answered a question. The third area is a combination of the first two, speech analysis followed by speech synthesis. This has application in secure voice transmission and speech data rate reduction. As an example of the latter, the telephone company requires 64K bits/sec

to transmit speech. The Department of Defense standard for data rate reduction is 2.4K bits/sec. The Air Force is experimenting with data rates as low as 150 bits/sec which provides intelligible speech.

The advent of the general purpose digital computer in the mid-1960s provided speech researchers with a powerful tool. Numerous speech processing algorithms using digital signal process techniques have been developed for both analysis and synthesis. From using dynamic programming to time-warp speech prior to processing, to algorithms for extracting parameters to be used for speech synthesis, speech processing is a billion dollar a year business.

Various speaker-dependent word recognition systems are commercially available. These systems generally perform some type of spectral analysis on the incoming speech signal. The recognition process involves classical pattern recognition techniques. These systems have a very high rate of successful recognition.

The success of these systems notwithstanding, the problem of constructing a speaker-independent recognition system remains unsolved. The solution to this problem involves determining what features of speech contain the information and hence are speaker independent. Before one can talk about extracting the information content from the

speech signal, a look at a model of how humans produce speech is in order.

A. FUNDAMENTALS OF SPEECH

Flanagan [Refs. 1 and 2] formulated a generally accepted model for human speech production. His model describes the vocal tract as a nonuniform acoustic tube connecting the vocal cords and the lips. In an adult male the vocal tract is approximately 17 cm. in length.

The vocal tract can be connected to an ancillary cavity called the nasal cavity. The coupling is accomplished through a trapdoor mechanism called the velum. The nasal cavity begins at the velum and terminates at the nostrils. In an adult it is about 12 cm. long. When non-nasal sounds are produced the velum closes, thereby sealing off the nasal cavity.

Humans are capable of producing two types of sounds, voiced and unvoiced. In the case of voiced sounds air moves over the vocal cords causing them to vibrate in a quasi-periodic fashion. Unvoiced sounds are generated by either forming a constriction in the tract and forcing the air through at high velocity or by allowing pressure to build up behind the closure and then releasing it suddenly. The name fricative is associated with the former while plosive is the name given to the latter.

Since the physical configuration of the vocal tract changes with time, Flanagan's model can be represented as a linear time-varying system as shown in Figure 1.1.

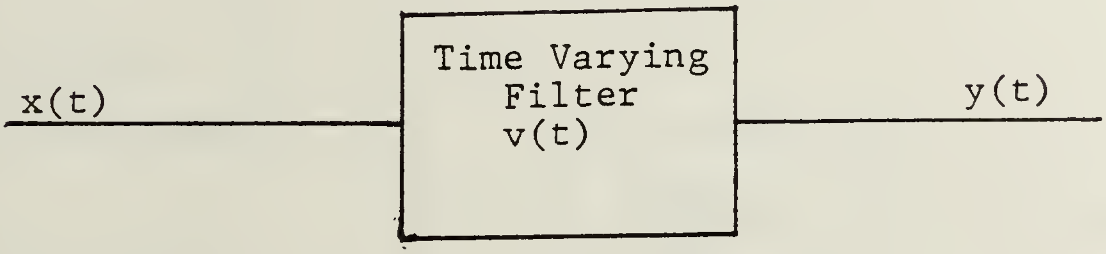

Figure 1.1. Model of Speech Production

If it is assumed that the vocal tract changes slowly with time the output can be approximated by the short-term convolution of the excitation, x(t), and the vocal tract impulse response, v(t). For voiced sounds x(t) is quasiperiodic hence the output y(t) is also quasiperiodic. For the unvoiced case the excitation x(t) is random and is generally approximated by white noise.

If the vocal tract impulse response of an individual could be obtained, then using the time varying linear system model intelligible speech should be able to be generated. The excitation would either be periodic or random depending on whether voiced or unvoiced sounds are

desired. Figure 1.2 is a simplified speech synthesis machine where the vocal tract parameters are stored in the RAM and downloaded to the voice synthesis chip which is excited by either the periodic or the random signal. This type of speech synthesis arrangement is the basis for Texas Instruments' (TI) Speak and Spell toys. TI can custom manufacture a speech synthesis chip which will emulate anyone's voice for $15,000.

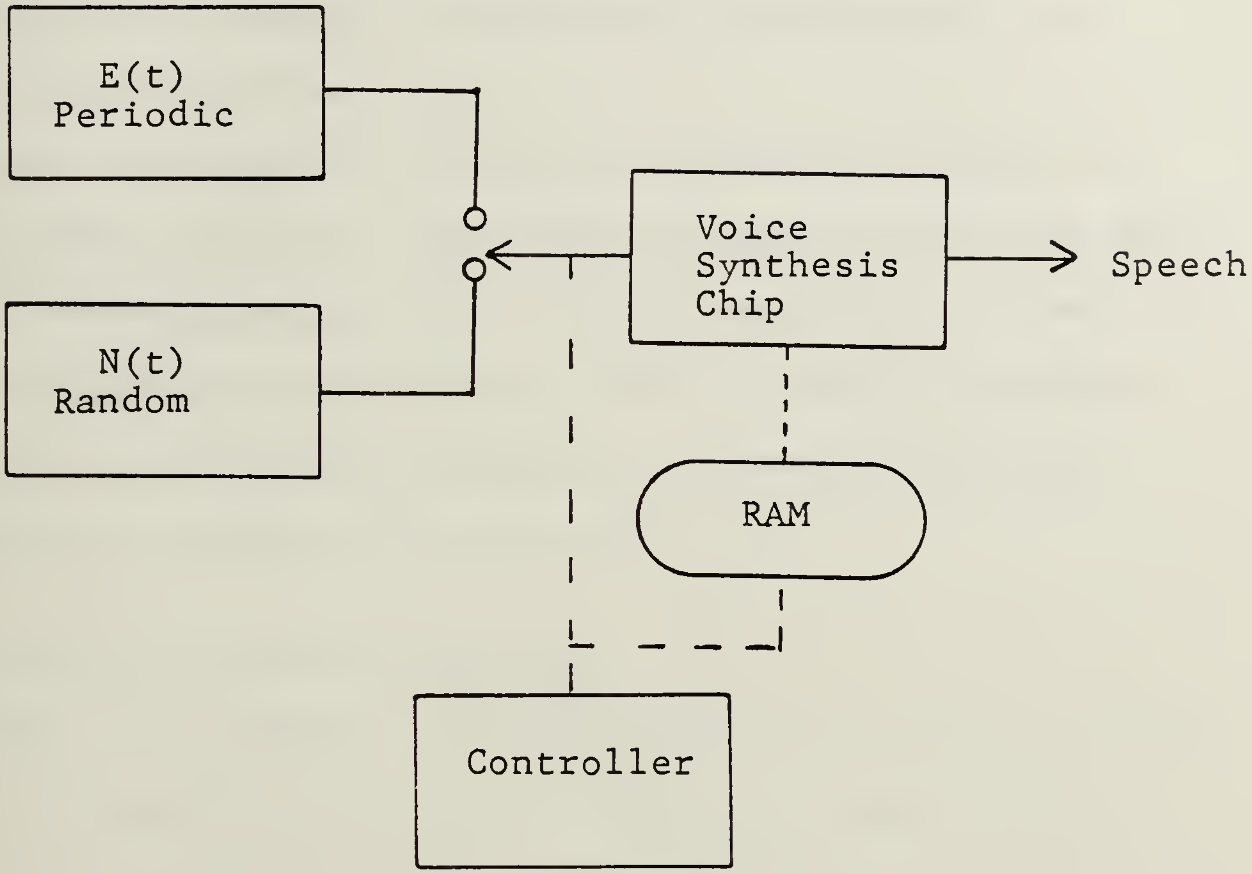

Figure 1.2. Voice Synthesis

These voiced and unvoiced sounds are combined in a unique fashion to form phonemes, the basic building blocks of language. All languages can be reduced to a finite number of these distinguishable building blocks. Phonemes are of such fundamental importance that if one phoneme is exchanged for another the meaning of an utterance is completely altered.

Thus, in theory, if a machine could be designed to disassemble utterances into their phoneme components the speech recognition problem would be completely solved. Despite vast amounts of time, effort, and money expended, however, the phoneme disassembler is years away from becoming an appears to be reality.

B. SPEECH RECOGNITION MACHINES

While the phoneme disassembler does not exist, several types of speech recognition systems are commercially available. The majority of these systems are classified as isolated word recognizers. As the name implies the systems are designed to recognize isolated words. The vocabulary of these machines is usually limited to 100-300 words and these systems are extremely speaker dependent. Thus, a person desiring to use these machines must first train the machine to recognize his voice. During the training phase the speaker's utterances are processed and templates formed. The recognition process involves comparing the incoming

utterance with those templates stored in the machine's memory [Ref. 3]. Although these machines have a limited vocabulary and cannot recognize connected or conversational speech, they are extremely useful for inventory control, quality assurance control, or for a pilot to check the systems in a combat aircraft. In all these instances the vocabulary is limited, the speaker is known, and voice data entry frees the individual to perform other tasks.

ITT has developed a word recognition system for the Air Force's F-16 fighter. The system is capable of recognizing 300 words and allows the pilot to check the status of certain systems while he maintains two hand control of the plane. This two-hand control is particularly important during low level, high speed attack runs. The pilots up-date their voice patterns monthly or if their voice changes due, say, to a cold. The patterns are stored in a bubble memory and inserted into the system prior to take-off. The microphone is located inside the pilot's oxygen mask and the system status is displayed on the cockpit's CRT. At a recent demonstration of this system it had a correct recognition rate of 99%.

The NPS Speech Processing Laboratory acquired an isolated word recognition system for experimentation purposes. The system is the VRM Voterm-2 manufactured by Interstate Electronics Corporation. The system, acquired in 1981,

weighs 10 lbs. and cost $2500. Today the same system has been reduced to a four chip set, for a cost of $1000.

The operation of the VRM is typical of the word recognition systems currently available [Ref. 4]. It allows the user to select the vocabulary size, decision threshold and number of training passes. It also allows for reference pattern transfer between itself and the host computer. The host computer serves only as a mass storage device and controller. All processing and recognition is performed real-time by the VRM.

The input speech signal is analyzed by a 16-filter analog spectrum analyzer and then passed through an A/D converter. This digitized speech data is then converted to a fixed-size (120 bit) pattern that preserves the information content of the utterance. During the training phase the VRM rejects utterances that do not sufficiently agree with previous training samples of the word. This rejection leads to a reduction of the number of 'ones' stored in the pattern. After seven training passes the pattern contains approximately one hundred 'zeroes'.

In 1980, NATO and the Rome Air Development Center (RADC) [Ref. 5] conducted a comparison test on three isolated word recognition systems. The vocabulary used consisted of the ten single digits of the respective languages of the speakers. The machines evaluated were the VRM system, the

Threshold Technology 8040 Preprocessor (cost $50,000) and the Nippon Electric DP-100 (cost $60,000).

Table 1.1 lists the results from the RADC test [Ref. 6]. Each speaker trained the machines by repeating each digit ten times. No attempt was made to introduce speakers who had not trained the machine. However, tests run at the Speech Processing Lab with the VRM with some non-trained speakers, using the ten digits and three sets of reference patterns the successful recognition rate for new speakers was less than 30%.

Thus, these systems work extremely well for what they were designed to accomplish. As previously stated, the basic question of what parameters of speech are speaker independent still remains unanswered. Numerous theories have been proposed and all have been unsuccessful. There is a lack of understanding of the human mechanisms used in understanding speech.

TABLE 1.1

RECOGNITION PERCENTAGES FOR RADC/NATO TEST

Language Spoken	Native or Non-Native	#Speakers	VRM Rec%	Thresh Rec%	VRM Thresh #Utters	NEC Rec%	NEC #Utters
English	Native	8	98.75	99.02	5400	99.64	3600
English	Non	9	98.78	98.69	4500	100.00	3000
French	Native	4	99.22	99.28	1800	99.67	1200
French	Non	1	93.22	95.44	900	99.83	600
Dutch	Native	3	96.20	98.13	1500	100.00	1000
All	Native	19	98.09	98.64	9900	99.74	6600
All	Non	10	97.85	98.15	5400	99.97	3600
All/Male	All	21	98.34	98.41	12300	99.86	8200
All/Female	All	8	96.60	98.70	3000	99.65	2000

II. MODELS OF THE EAR

For a long time people have been trying to understand how the human ear functions. In the first century B.C., the Roman poet, philosopher Lucretius postulated a model "involving little grains of sand in the inner ear responding too different tones" [Ref. 7]. The 18th century Italian violinist Tartini noted that the ear produced a third tone from two tones played simultaneously. Thus the long held belief that the ear was a linear device was demonstrated to be false. Today the ear is thought to be a nonlinear device even at power levels near the threshold of hearing.

The first concentrated research into the process of hearing did not begin until the mid-1800's. This was the time of Seebeck, Helmholtz, and Ohm. It was Ohm who postulated a now famous law on the relationship of speech and its phase angle. He stated that all the information content of speech is contained in its power spectrum and was independent of the phase angle of the components. Although Ohm's law has been modified in recent years, it remains as one of the fundamental laws of psychoacoustics.

The ear can be broken down into three physical areas; the outer, middle and inner ear. Sound waves impinge on the outer ear and are conducted down a canal until they reach

the middle ear. The middle ear contains three tiny bones. The alternate compressions and refractions of the speech wave cause the eardrum to strike the bones. In the inner ear the wave travels along a thin membrane whose frequency response varies between 100 Hz and 20 KHz. This provides for spectral analysis of the incoming signal.

The membrane of the inner ear is lined with tiny hairs. It is these hairs or more correctly groups of hairs that perform the spectral analysis. Recent studies at the California Institute of Technology [Ref. 8] have found that each tiny hair bundle consists of 30-150 thin, rod-shaped extensions called cilia. These hair bundles are attached to hair cells. The hair cells are very sensitive transducers which convert the movement of the hair bundle into an electrical signal which is sent to the brain. The hair bundle-hair cell combination form a sort of mechanical spectrum analyzer.

Manfred Schroeder [Ref. 9] describes an experiment in which the inner ear's sensitivity to phase was demonstrated. The experiment was as follows:

1) A 100 sec. sample of speech was Fourier transformed.
2) Random phase angles were assigned to the frequency components (assuming a uniform distribution 0 to 2π).
3) The inverse Fourier transform was taken.

The resultant signal sounded like white noise. Thus by randomizing the phase angles the signal was transformed from

intelligent speech to noise. This lent credence to the hypothesis that the inner ear was phase sensitive and that Ohm's law, if not wrong, was at least in need of modification. The experiment was repeated this time using a 50 msec. sample of speech. The resultant signal was non-intelligible noise. Ohm's law modified to say that only the short term amplitude spectrum contained the speech information appeared to be correct.

Ohm based his law on a model of the ear that said:

1) The ear has a tuned bandpass filter covering the audio range.
2) Only the output amplitude of each filter is sent to the brain.

Today the most likely candidate for the bandpass filter are the hair bundle-hair cell combinations that respond to only selected stimuli.

In 1947 an experiment was conducted [Ref. 10] in an effort to obtain a definite answer to the phase sensitive question. An AM signal at 2000 Hz was modulated by a 100 Hz signal. Thus three frequency components (1900 Hz, 2000 Hz, 2100 Hz) were present. One of the sidebands had its phase shifted by 180°. This phase shift resulted in what was termed a quasi-FM (QFM) signal. Upon listening to the signals there was a noticeable difference between the AM and QFM signal. Thus there was a revived interest in the ear's

capability to discern waveforms and not just their amplitude.

In a further effort to determine to what extent phase is important in discerning speech, Hall and Schroder [Ref. 11] conducted an experiment where the phase angle of one of two pure tones was changed. Specifically two tones one at 200 Hz and 0° and another at 400 Hz but with phase angles of 0°, 60°, 120°, 180°, 240°, and 300° were listened to, three signals at a time. The listeners' task was to determine which two signals sounded most alike and which two sounded least alike. The results showed that those harmonics of 400 Hz whose phase angle differed the least were judged to be the most similar consistently.

About twelve years prior to this experiment researchers at Bell Labs postulated that the phase dependency seen in experiments involving the inner ear could be traced to the phase dependence of the inner and middle ear distortion products. Due to the presence of these nonlinear distortion products a new spectrum, called the inner spectrum was formed in the inner ear. It is this spectrum that is analyzed by the hair bundles of the inner ear.

This theory certainly would explain what happened at Bell Labs during a 1958 experiment [Ref. 12]. When the phase of one of 31-equal amplitude harmonics all 0° phase was changed to a 180° a pure tone was heard. This tone was

not heard when the signal was put through a loud speaker. Thus using the inner spectrum theory changing the phase of one harmonic to 180° altered the amplitude of one of the distortion products. This altered the inner spectrum causing a bump in the spectrum where previously it had been flat.

In Germany, Terhardt and Fastl [Ref. 13] conducted experiments trying to connect frequency difference and phase angles. They formed a signal $s(t) = a_1 \cos(2\pi f_1 t) + a_2 \cos(2\pi f_2 t - \phi_2)$ where $f_1 = 200$ Hz, $f_2 = 400$ Hz and asked listeners to adjust the amplitude of each component so the 400 Hz tone was just audible. This was to be done while the phase angle, ϕ_2, of the 400 Hz tone was changed. The results showed that when ϕ_2 was changed from 0° to 180°, the amplitude of the 400 Hz signal had to be increased by 12 dB to remain audible.

Yet another theory on the functioning of the ear came out of this experiment. The researchers theorized that the hair cells of the ear were discerning the time between successive spikes in the waveform and passed this information to the brain. This appeared as a reasonable explanation as when $\phi_2 = 0^\circ$ the time between successive spikes was 2.5 msec. With $\phi_2 = 180^\circ$ the time between spikes was 5 msec., unless the amplitude of the 400 Hz tone was increased by considerable

amount. With the amplitude increased the small spikes at the 2.5 msec. mark would increase dramatically.

This theory is consistent with the physiology of the ear. All the electric pulses transmitted to the brain from the hair cells have approximately the same amplitude, thus the timing between the pulses is the information that they carry.

From the myriad of theories presented it is easy to conclude that a definitive model of the human ear is non-existent. The fact that phase contains some information content has been demonstrated. Whether phase alone is the speaker independent feature that researchers are looking for remains an unanswered question. Experiments conducted in the late 1970's and 1980's using phase-only representations of speech have given some creditability to the hypothesis that phase must be included as one of the speaker independent features of speech.

III. PHASE-ONLY REPRESENTATIONS OF SPEECH

Recapitulating, Ohm's law stated that all the information content of speech could be obtained from the short term power spectrum and that phase angle of the components was meaningless. Thus, in the short term the ear is phase deaf. Oppenheim [Ref. 14] sought to explore more fully the importance of phase in speech.

Given the Fourier transform of a speech signal

$$F(\omega) = |F(\omega)|e^{j\theta(\omega)} \tag{3.1}$$

and if the $|F(\omega)|$ is set equal to one, the inverse transform of $e^{j\theta(\omega)}$ is a phase only representation of the speech. This phase only representation retained total intelligibility, while exhibiting the characteristics of being high passed filtered and having white noise added. The magnitude only representation was speech-like in its appearance but was not intelligible.

Oppenheim concluded that transforming a signal to its phase only form was equivalent to passing it through a spectral whitening process with a filter whose response is $H(x) = 1/|F(x)|$, where $F(x)$ is the Fourier transform of the original signal. This spectral whitening did not destroy the intelligibility of the speech.

Contrary to Ohm's law, Cox and Robinson [Ref. 15] conducted a series of four experiments which preserve the short term phase of a speech signal while either destroying or severely distorting the amplitude. These phase-only signals were found to retain many speech characteristics and were intelligible to the listeners. Hence under certain transformations short term phase may be one of the physical invariants of speech.

The experiments used a speech signal that was analog band limited to 8 KHz and sampled at a rate of 20 KHz with 12 bits A/D. Successive 25.6 msec windows, corresponding to 512 data points, were fast Fourier transformed. Nonlinear operations were applied to each data set, and the inverse fast Fourier transforms were taken yielding 25.6 msec of reconstructed speech signal. These signals were D/A converted at a rate of 20 KHz and passed through a 8 KHz low pass analog filter. Only rectangular windows were used and no attempt was made to fit the windows together since amplitude of the reconstructed signal was umimportant. The first two experiments are included for completeness only. The latter two are the concern of this thesis.

A. SHORT-TERM PHASE ONLY SIGNALS

This experiment basically repeated the previously mentioned work of Oppenheim, as the magnitude of the Fourier transform of the data sets was set equal to one. The phase

was unchanged. The reconstructed short-term phase only signal was found to retain many of the original waveform's features. Listeners could identify speaker dependent characteristics and the intelligibility, while not judged good, was likened to a signal containing a lot of noise. There was no attempt made by the researchers to clean up the signal. The results of this experiment clearly are contrary to Ohm's law and demonstrate that short-term phase only speech is intelligible.

B. ANALYTIC SIGNAL PROCESSING

The second experiment was a repeat of one carried out in the late 1940's. Here the representation is an infinitely clipped version of the original signal

$$Sc(t) = \mathrm{Sgn}\,[s(t)] \tag{3.2}$$

where s(t) is the original signal, and Sgn is defined to be the sign of s(t). Thus the continous valued signal, s(t), was transformed into a discrete valued signal. The transformation retains only the real-zero information of s(t). That is, if s(t) was an analytic signal the real-zeros mark the time when the phase was changed by 180°. The intelligibility of such a signal was not commented on by the experimenters, however, they did say that large amounts of speech information were retained using this transform.

C. DIRECT PHASE CEPSTRUM

The concept of cepstral analysis of speech was developed by Oppenheim [Ref. 16] and is an example of a broad class of nonlinear processing called homomorphic processing. These homomorphic systems obey generalized laws of superposition. If $x_1(n)$ and $x_2(n)$ are inputs to a homomorphic system and $y_1(n)$, $y_2(n)$ are corresponding outputs and k is any scalar then

$$y_1(n) = \phi[x_1(n)]$$

$$y_2(n) = \phi[x_2(n)]$$

$$\phi[x_1(n) \,\Delta\, x_2(n)] = \phi[x_1(n)] \,\square\, \phi[x_2(n)]$$

$$\phi[k \bigcirc x_1(n)] = k * y_1(n)$$

where Δ, $\square$, $\bigcirc$, and $*$ are mathematical operations.

The importance of these homomorphic systems is that ϕ can be broken down into a cascade of operations as shown in Figure 3.1 where A_o, A_o^{-1} are inverses of each other and L is a simple linear filter.

Thus Oppenheim [Ref. 17] formulated a model for the production of speech as shown in Figure 3.2. The model is based on the assumption that the excitation and vocal tract parameters are independent. The source of excitation for the voiced sounds is the impulse generator whose period is

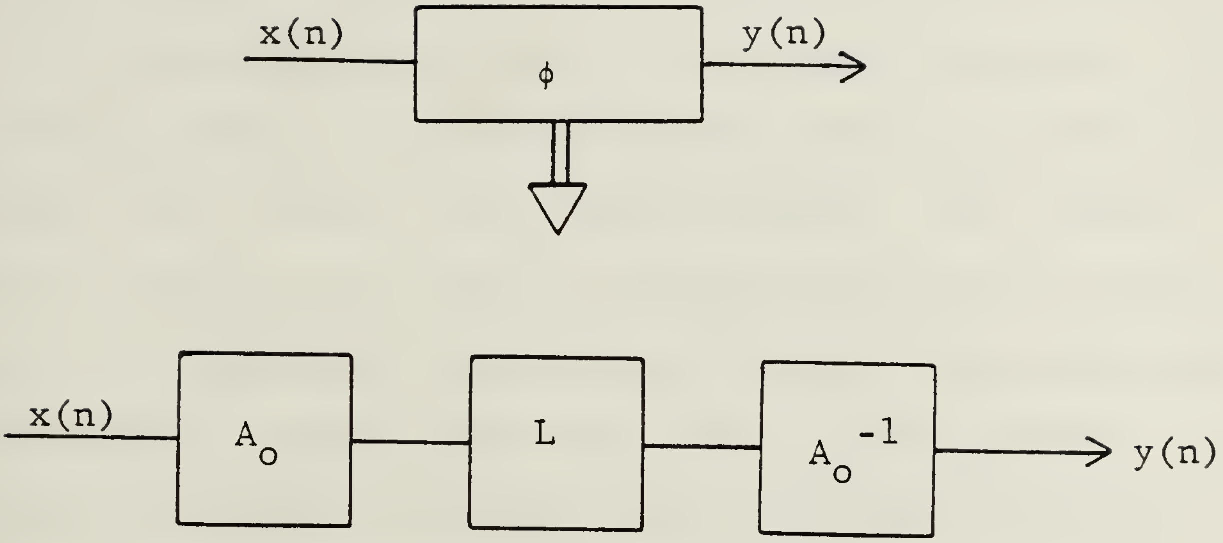

Figure 3.1. Homomorphic System

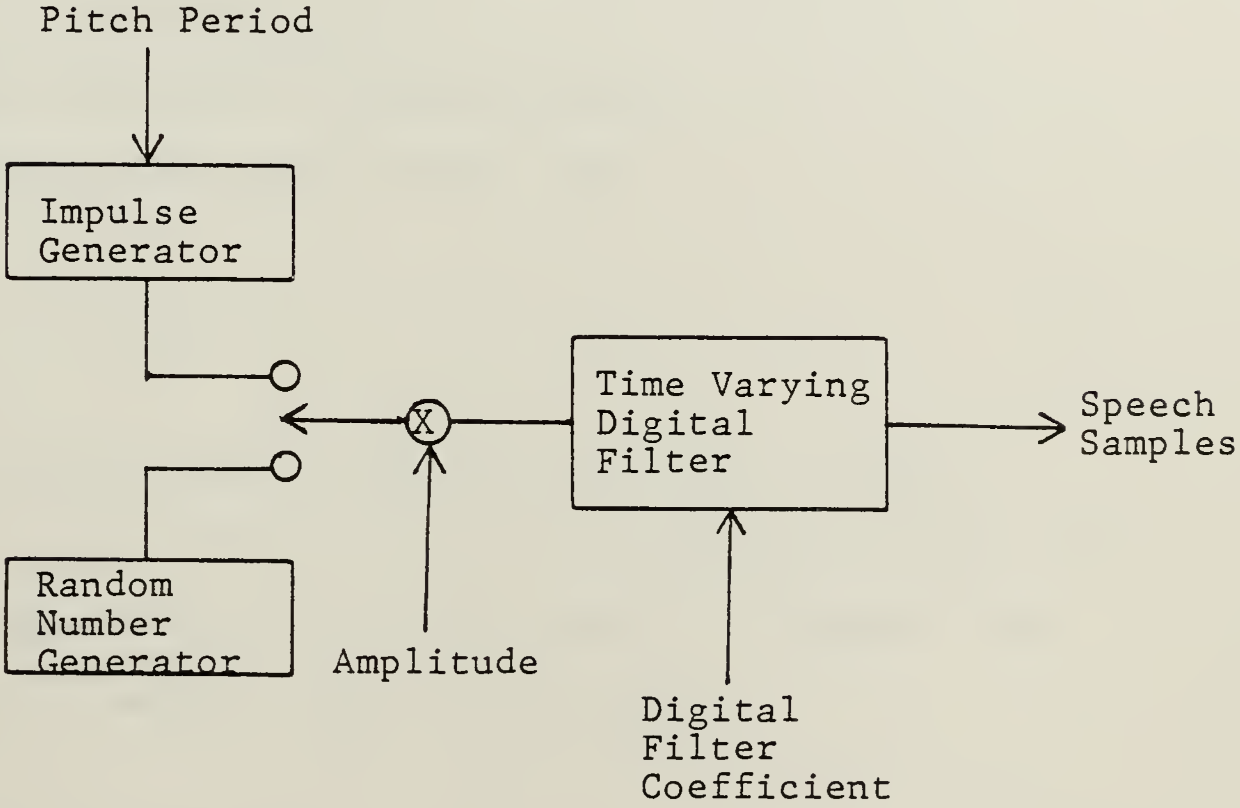

Figure 3.2. Model for Speech Production

controlled by the pitch-period signal. The impulse generator produces an impulse once every N_o samples, where N_o is the pitch-period and $1/N_o$ is the pitch frequency. The unvoiced excitation is from the random number generator and simulates both fricative and plosive sounds. The digital filter is assumed to be slowly varying with time and hence changes its coefficients once every 10 msec. The amplitude control simply adjusts the output level of the speech.

Using this model the output digitized speech waveform consists of the convolution of

(1) The train of impulses representing the pitch

(2) The excitation pulse

(3) The vocal tract impulse reponse.

If x(n) denotes the output signal, then

$$x(n) = [p(n) * e(n) * u(n)]\, w(n) \tag{3.3}$$

where p(n) is the train of pitch pulses, e(n) is the excitation pulse, u(n) the vocal tract impulse response, and w(n) the window through which the speech is viewed. The window w(n) is smooth, hence we can define

$$\hat{p}(n) = p(n)\, w(n) \tag{3.4}$$

Then substituting this into equation (3.3) it is possible to approximate x(n) by

$$x(n) \approx \hat{p}(n) * e(n) * u(n) \qquad (3.5)$$

Examining equation (3.5) it is possible to convert the triple convolution into a triple sum by first taking the Fourier transform and then taking the logarithm. Processing of this signal can be accomplished by a linear system and recovery of the waveform can be made by passing the processed signal through an exponentator followed by inverse Fourier transformer. Thus a homomorphic system for processing speech has been developed, as shown in Figure 3.3 [Ref. 18].

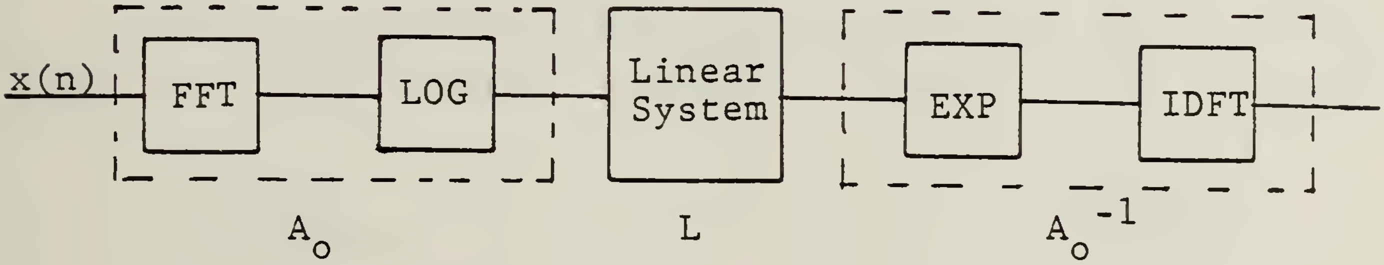

Figure 3.3. Homomorphic System for Processing Speech

Variations on this basic system have been developed to estimate parameters of both the vocal tract transmission

functions and the excitation functions. One of these variations involves making the assumption that the excitation is $s(n) = \hat{p}(n) * e(n)$, then equation (3.5) can be written as

$$x(n) = u(n) * s(n) \tag{3.6}$$

The system to process signals given by equation (3.6) is shown in Figure 3.4 [Ref. 19].

Referring to Figure 3.4, the signal at A is x(n) and the signal at D is called the cepstrum of x(n) and equals the cepstra of the excitation plus the cepstra of the vocal tract impulse response.

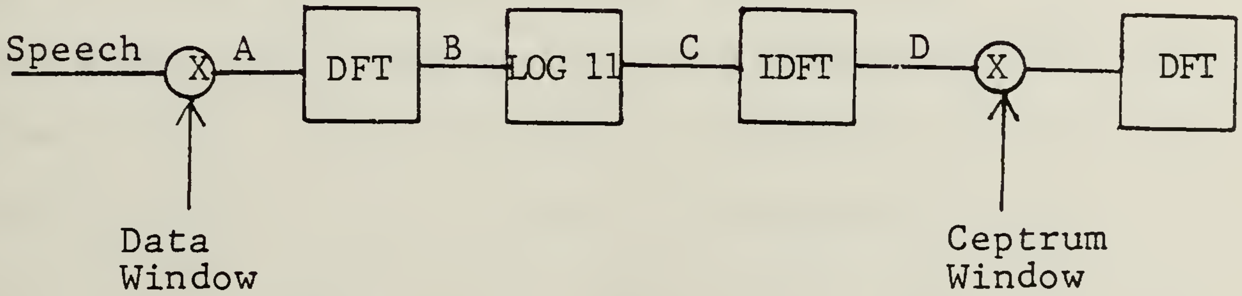

Figure 3.4. Cepstral Processing of Speech

An important feature of the cepstrum at D is that it separates the excitation from the vocal tract response. The excitation is a sequence of quasi-periodic pulses, thus its

Fourier transform, at point B, is a line spectra where the lines are spaced at harmonics of the fundamental frequency. The log magnitude operation does not effect the general shape of the spectra. The IDFT of the signal produces another quasi-periodic waveform with pulses spaced at the fundamental period. Thus the cepstrum of the excitation should consist of pulses around $n = 0, T, 2T, \ldots,$ where T is the pitch period.

The DFT of the vocal tract response is a slowly varying function of frequency. The log magnitude and IDFT yield a sequence that is negligible after a few samples. The cepstrum at D consists of two sequences, one which is negligible after a few samples and one that is periodic. Thus the cepstrum at D does differentiate the excitation from the vocal tract parameters. The use of the cepstral processing has been extended into many diverse fields [Ref. 20].

For their third experiment, Cox and Robinson [Ref. 21] modified Figure 3.4 by setting the magnitude of the signal at point C equal to one. Hence the cepstrum at point D is due only to the phase of the signal at A. What amount of information and intelligibility does this phase only cepstrum contain? Surprisingly the cepstrum was judged to be very intelligible by listeners and the noise level was reduced when compared with the short-term phase only speech (experiment number one).

D. INSTANTANEOUS PHASE OF THE ANALYTIC SIGNAL

The fourth experiment performed by Cox and Robinson [Ref. 22] was first performed in 1955 by Marcoui and Daguet who were looking for more efficient modulation techniques. They sought to use the analytic signal representation of a real signal s(t). Given a real signal s(t), which is Hilbert transformable, form a quadrature signal $s^*(t)$ and construct

$$m(t) = s(t) + j\, s^*(t) \tag{3.7}$$

From equation (3.7) it is possible to recover the original signal as

$$s(t) = \mathrm{RE}[m(t)] = |m(t)| \cos \theta(t) \tag{3.8}$$

Equation 3.8 lets the real signal, s(t), be represented by a magnitude and phase.

The concept of an analytic signal, which equation (3.7) is called, was meaningless for discrete-time signals, until Rabiner and Schafer [Ref. 23] developed a complex representation for real discrete-time bandpass signals.

Following the notation of Rabiner and Schafer, given a real sequence, x(n), with Fourier transform X(w), construct a complex sequence

$$\tilde{x}(n) = x(n) + j\,\hat{x}(n) \tag{3.9}$$

The Fourier transform of which is

$$\begin{aligned} \tilde{X}(\omega) &= 2\,X(\omega) && 0 \leq \omega < \pi \\ &= 0 && \pi \leq \omega < 2\pi \end{aligned} \tag{3.10}$$

From equation (3.9) the Fourier transform of x(n) is

$$\tilde{X}(\omega) = X(\omega) + j\,\hat{X}(\omega) \tag{3.11}$$

and from equation (3.10) it follows that

$$X(\omega) + j\,\hat{X}(\omega) = 0 \qquad \pi \leq \omega < 2\pi$$

and

$$\tilde{X}(\omega) = 2X(\omega) \qquad 0 \leq \omega < \pi$$

These requirements are satisfied if

$$\hat{X}(\omega) = H_d(\omega)\,X(\omega) \tag{3.12}$$

where

$$\begin{aligned} H_d(\omega) &= -j && 0 \leq \omega < \pi \\ &= +j && \pi \leq \omega < 2\pi \end{aligned} \tag{3.13}$$

Thus given any sequence x(n), it is possible to obtain the sequence $\hat{x}(n)$ by linear filtering of x(n) with a filter whose frequency response is given by equation (3.13). Such a filter is called an ideal Hilbert transformer and $\hat{x}(n)$ is the Hilbert transform of x(n). The impulse response of the ideal Hilbert transformer is

$$h_d(n) = \frac{2}{\pi} \frac{\sin^2 (\frac{\pi n}{2})}{n} \qquad n \neq 0 \qquad (3.14)$$

$$= 0 \qquad n = 0$$

Examining equation (3.14), the impulse response is non-causal, of infinite duration, has odd symmetry, and all even-numbered samples are equal to zero (i.e., $h_d(2n) = 0$, $n = 0, \pm 1, \pm 2, \pm 3, \ldots$).

Since infinite length, non-causal impulse responses are not realizable an FIR approximation is required. Given a causal FIR system whose impulse response is h(n), $0 \leq n \leq N-1$, its frequency response is given by

$$H(\omega) = \sum_{n=0}^{N-1} h(n) e^{-j\omega n} \qquad (3.15)$$

Equation (3.13) says the desired frequency response, $H_d(\omega)$, is purely imaginary. Thus the real part of equation (3.15) must equal zero as h(n) is real. In order for the real part of equation (3.15) to be zero h(n) must satisfy the symmetry condition

$$h(n) = -h(N-1-n) \qquad n = 0, \ldots, N-1. \tag{3.16}$$

If N is odd, h(n) has odd symmetry about $n = (N-1)/2$. If N is even, h(n) has odd symmetry about a point halfway between the samples at $n = N/2$ and $n = (N/2) + 1$. If equation (3.16) is satisfied, equation (3.15) can be written as

$$H(\omega) = e^{-j\omega(N-1)/2} \, [jH^{*}(\omega)] \tag{3.17}$$

where $H^{*}(\omega)$ is a real function of ω. If N is odd, $H^{*}(\omega)$ can be written as

$$H^{*}(\omega) = \sum_{n=1}^{(N-1)/2} a(n) \sin(\omega n) \tag{3.18}$$

where $a(n) = 2h\left(\frac{N-1}{2} - n\right), \ n = 1, 2, \ldots, \left(\frac{N-1}{2}\right)$ (3.19)

Also for N odd,

$$h\left(\frac{N-1}{2}\right) = 0 \tag{3.20}$$

For N even, equation (3.18) becomes

$$H^{*}(\omega) = \sum_{n=1}^{N/2} b(n) \sin[\omega(n - 1/2) \tag{3.21}$$

where $b(n) = 2h\left(\frac{N}{2} - n\right), \ n = 1, \ldots, N/2$

Examining equation (3.17) more closely, we find that the factor $e^{-j\omega(N-1)/2}$ is a delay of (N-1)/2 samples.

In finding an approximation to the ideal Hilbert transform, coefficients a(n) and b(n) were chosen in such a fashion that $jH^*(\omega)$ approximates the ideal frequency response given by equation (3.13). Thus $H^*(\omega)$ must approximate

$$\begin{aligned} D(\omega) &= -1 \qquad 2\pi F_L \leq \omega \leq 2\pi F_H \\ &= +1 \qquad 2\pi(1 - F_H) \leq \omega \leq 2\pi F_L \end{aligned} \tag{3.22}$$

where F_L and F_H are the lower and upper cutoff frequencies represented as fractions of 2π. From equation (3.18), $H^*(\omega)$ must equal zero at $\omega = 0$ and $\omega = \pi$ when N is odd and must equal zero at $\omega = 0$ for the case when N is even.

For the ideal transformer the impulse response was zero for all even numbered samples and the frequency response was imaginary, odd, periodic and

$$H_d(\omega) = H_d(\pi - \omega).$$

For the FIR approximation similar properties must be valid. If N is odd and $F_L = .5 - F_H$ and assuming that

$$H^*(\omega) = H^*(\pi - \omega). \tag{3.23}$$

Then substituting into equation (3.18) yields,

$$\sum_{n=1}^{(N-1)/2} a(n)\ \sin(n\omega) = \sum_{n=1}^{(N-1)/2} a(n)\ \sin[(\pi - \omega)n]$$

$$= \sum_{n=1}^{(N-1)/2} a(n)\ (-1)^{n+1} \sin(\omega n)$$

rearranging terms

$$\sum_{n=1}^{(N-1)/2} a(n)\ \sin[\omega n(1 - (-1)^{n+1})] = 0$$

Thus $a(n) = 0$ $\quad$ n even

$= \text{unconstrained}$ $\quad$ n odd.

Combining this result with equations (3.16), (3.19), and (3.20) have that for $(N-1)/2$ even, $h(n) = 0$, for $n = 0, 2, \ldots$ and when $(N-1)/2$ is odd, $h(n) = 0$, for $n = 1, 3, 5, \ldots$. For the case of N even no relationship among the coefficients exist.

One important difference between even and odd length impulse responses can be seen in direct convolution. The convolution summation given by

$$\hat{x}(n) = \sum_{k=0}^{N-1} h(k)\ x(n-k)$$

involves only $(N+1)/4$ multiples per output sample for N odd and $N/2$ multiples for N even. The saving occurs because

alternate values of h(n) are zero for N odd. Because of this savings and for technical considerations only Hilbert transformers of odd length are used.

In determining the values of h(n), Rabiner and Schafer [Ref. 24] used the Remez algorithm for the design of optimal FIR filters. The values of h(n) were calculated to minimize the peak approximation error which is given by

$$G = \text{MAX}\ [D(\omega) - H^*(\omega)] \qquad (3.24)$$

$$2\pi F_L \leq \omega \leq 2\pi F_H$$

The Remex algorithm gives a Chebyshev or equiripple approximation to the desired response. Hence the error function is equiripple over the range $2\pi F_L \leq \omega \leq 2\pi F_H$. Given an N, F_L and F_H the resulting approximation is best in the mimimax sense.

Using this concept of an analytic signal representation for discrete-time signals, Cox and Robinson [Ref. 25] formed the analytic phase representation of a speech signal. Given a sampled speech signal, s(n), they calculated the Hilbert transform, $s^*(n)$, by the use of a 79-weight Hilbert transformer. Thus having the analytic signal

$$m(n) = s(n) + j\, s^*(n)$$

the original signal s(n) is given by

$$s(n) = |m(n)| \cos \theta(n).$$

The analytic phase representation is given by cos $\theta(n)$. Thus by way of a mathematical artifice a real-valued sequence s(n) is represented as having magnitude and phase with the phase only being retained. Contrary to common sense, perhaps, this analytic phase representation of speech was found to be intelligible. While these experiments by themselves do not prove that phase is a physical invariant of speech, they do indicate that more research is needed to determine to what role phase plays in speech intelligibility.

As was mentioned, a 79-weight Hilbert transformer was used in obtaining the analytic signal. Rabiner and Schafer [Ref. 26] calculated weights for three different values of peak approximation errors and cutoff frequencies. Table 3.1 lists these weights and Figures 3.5 through 3.7 are plots of the magnitude of the frequency response. Table 3.1 only lists even weights, since 79 is odd, all odd weights are zero and the weights have odd symmetry about n = 39.

TABLE 3.1

HILBERT TRANSFORMER WEIGHTS

N = 79

	$F_L = .01$	$F_L = .02$	$F_L = .05$
n	G = .0388830	G = .0024390	G = .0000010
0	-.0229388	-.0019358	-.0000041
2	-.0075151	-.0017746	-.0000179
4	-.0087784	-.0025624	-.0000550
6	-.0101565	-.0035600	-.0001389
8	-.0117808	-.0048021	-.0003074
10	-.0135612	-.0063300	-.0006182
12	-.0155902	-.0081910	-.0011532
14	-.0179182	-.0104453	-.0020239
16	-.0206260	-.0131630	-.0033761
18	-.0237742	-.0164470	-.0053956
20	-.0274953	-.0204251	-.0083167
22	-.0319865	-.0252943	-.0124372
24	-.0375627	-.0313515	-.0181511
26	-.0447012	-.0390711	-.0260178
28	-.0542333	-.0492818	-.0369200
30	-.0677331	-.0635544	-.0524475
32	-.0885965	-.0852651	-.0759556
34	-.1256401	-.1232135	-.1161821
36	-.2111964	-.2097186	-.2053402
38	-.6362830	-.6357869	-.6343000

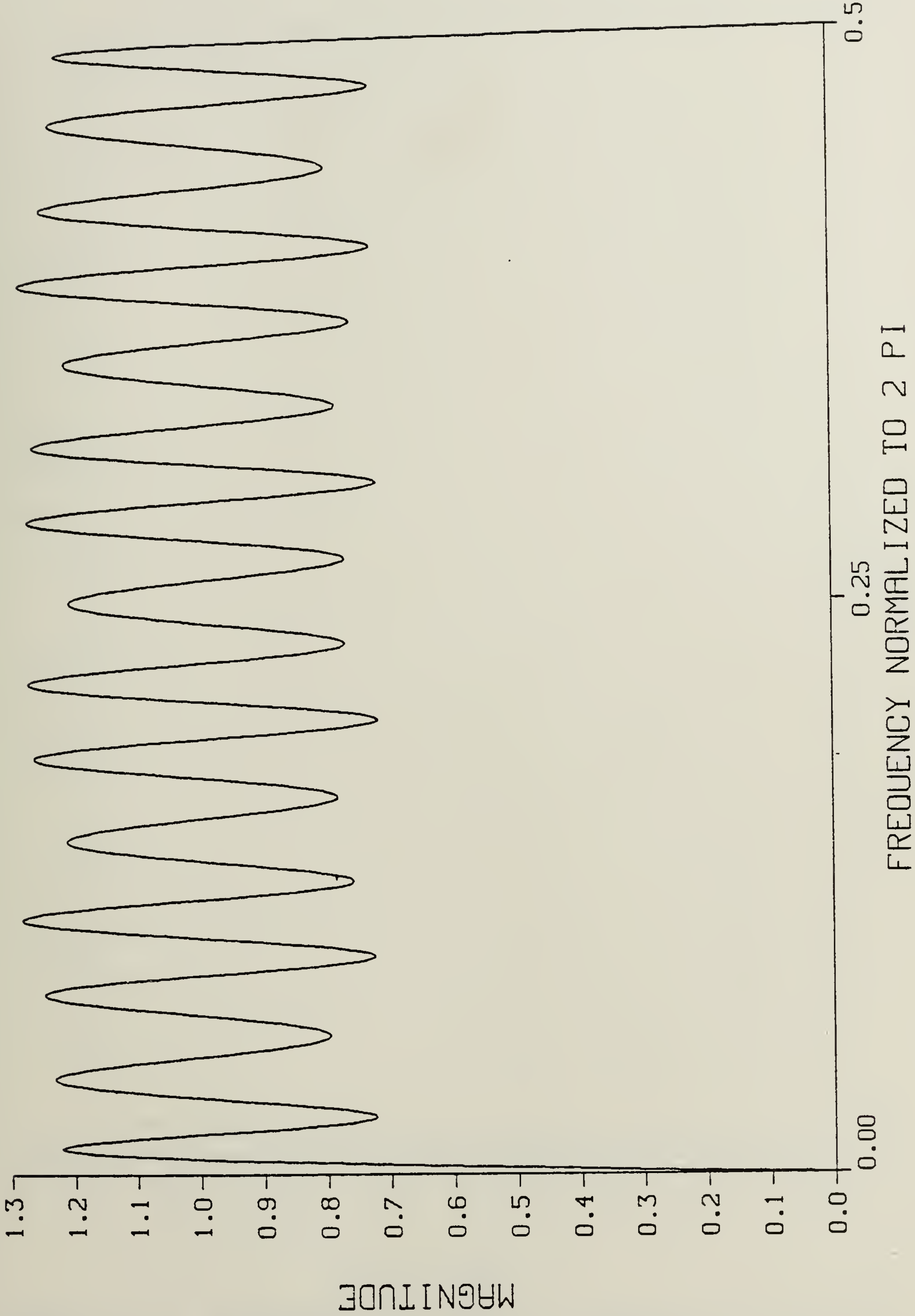

Figure 3.5. Frequency Response of Hilbert Transformer
$N = 79$, $F_L = .01$, $G = .0388830$

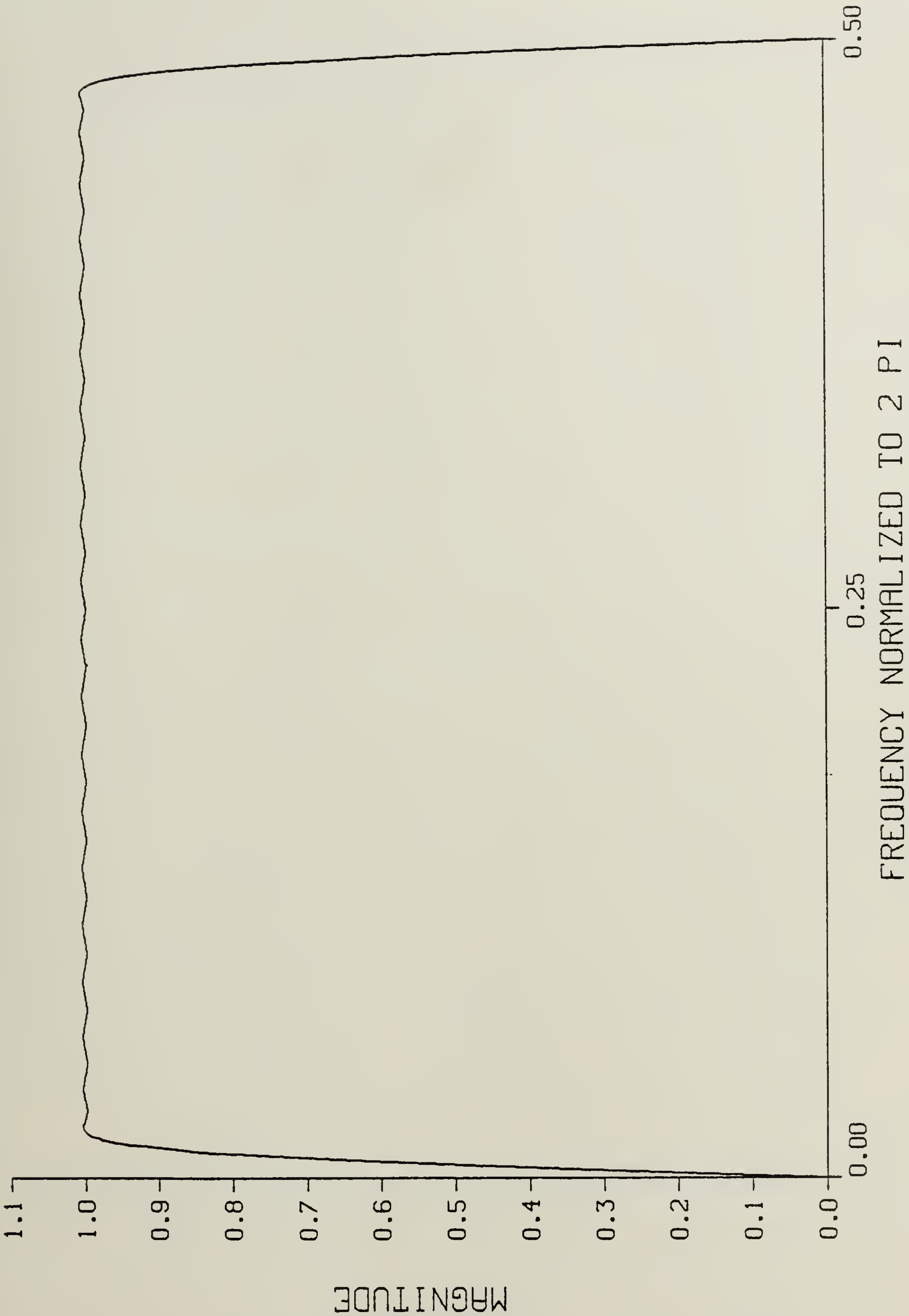

Figure 3.6. Frequency Response of Hilbert Transformer
$N = 79$, $F_L = .02$, $G = .0024390$

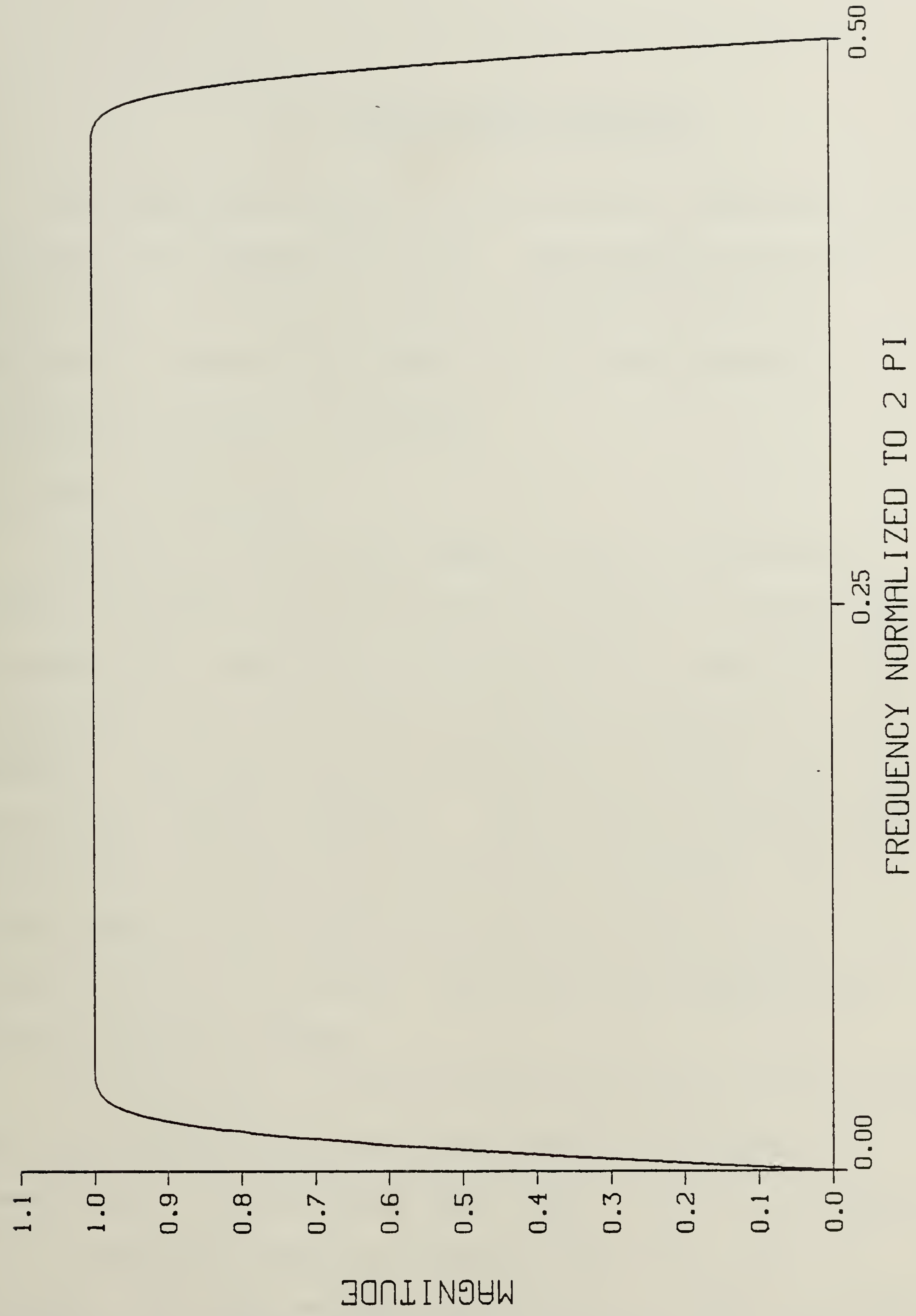

Figure 3.7. Frequency Response of Hilbert Transformer
$N = 79$, $F_L = .05$, $G = .0000010$

IV. EXPERIMENTAL PROCEDURE

This thesis extends the work of Cox and Robinson to the isolated word recognition field. Specifically using the homomorphic and analytic signal processing techniques employed in experiments three and four an isolated word recognition system is developed.

A. DATA ACQUISITION

In order to form a data base for use by the system twenty volunteers were recruited to record the digits zero through nine. Each participant was given a questionnaire/ instruction sheet like that contained in Appendix A. All speakers were males between the ages of 25 and 35 and all were native English speakers. Their places of birth varied from eastern Pennsylvania to southern Tennessee. Ten of these speakers were selected to form the data base or pattern base of the system. The other ten speakers were used to test the system.

The speech was recorded on an analog tape recorder with all recordings being done in the Speech Processing Laboratory. The recordings were done in the late afternoon or in the evening when the ambient noise level was at a minimum. The tape recorder used was the HP-3964A reel-to-reel instrumentation recorder running at 7.5 ips using AMPEX professional audio tape.

Before this analog speech could be digitized an appropriate bandwidth and sampling rate had to be determined. The power spectral density of each digit was computed and averaged over ten utterances of the digit. The majority of the power was found to be below 3 KHz except in the case of the number 'six' where nonnegligible power was found to frequencies up to 6 KHz. A cutoff frequency of 4 KHz was chosen, which is exactly half the bandwidth that Cox and Robinson used. As will be explained later, once the bandwidth is fixed the sampling rate is also fixed. In this case the sampling rate is fixed at 10 KHz.

The machine used to digitize the speech was the GENRAD 2505 Signal Analysis System [Ref. 27]. The system is a narrowband (0 - 25 KHz) signal analysis system originally designed for vibrational analysis studies. The system uses a DEC PDP 11/34A as the host computer and supports two channels of A/D conversion.

The heart of the system, softwarewise, is GENRAD's Time Series Language (TSL) which allows the operator to control the A/D converter. TSL is an interpretive language which uses commands similar to BASIC. The TSL program 'ANADSK' is the routine that provides analog input to disk storage. Given a bandwidth the 'ANADSK' routine sets the sampling rate at 2.56 times the highest frequency component to prevent aliasing. The system provides for high-speed

continuous sampling and writes the digitized data to the system's Winchester disks in 2048 byte blocks.

The two-channel A/D converter has two 6-pole Chebychev filters in cascade each with 96 dB/octave rolloff above cutoff per channel as anti-aliasing filters. The A/D converter is a 2 μsec converter with a 12 bit output.

Once the speech was digitized a time window for the sampled data had to be determined. Referring again to the utterances whose power spectral densities were computed, the average length of the utterances was 740 msec. In order for the mathematics to work out nicely a 750-msec window was chosen.

Using TSL library routines 'RTIO' and "XDISPL' a routine was written that displayed the digitized data on the system's CRT. The program graphically displayed 1024 samples at a time and allowed the operator to select any 256 samples for transfer to the W. R. Church Computer Center's IBM 3033 for processing. This transfer was via a 1200 baud modem. With the capability to view the data prior to transfer, the start of the utterance could be selected to within 128 samples. Since the time window was selected to be 750 msec and the speech was sampled at 10,240 samples/sec, 7680 points needed to be transferred. Thus thirty blocks of 256 samples each were transferred per utterance.

The transfer/interface program between the Speech Lab's PDP 11/34A and the IBM 3033 was written by LT Jay H. Benson. A copy of his program, 'CATCH', is included in Appendix B. The transfer of data via the modem was very time consuming as for technical reasons each sample which occupied two bytes on the PDP 11/34A was made into a four byte number for transfer. The sixteen most significant bits were then masked off prior to storage on the IBM system. In order to minimize the amount of disk storage required, the data was written to the disk using an unformatted FORTRAN write statement, using Integer * 2 numbers. Even using this scheme to maximize storage efficiency 24 cylinders plus magnetic tape backup were required to store the data.

B. DATA PROCESSING

The decision to use the IBM system to process the data was based on the availability of library routines (e.g., IMSL, NONIMSL), the DISSPLA graphics package, and the full screen text editor. All programs in Appendix B were written in FORTRAN H.

The first task was to compute an average waveform for the speaker. In order to accomplish this, three of the four utterances of each of 10 speakers were averaged together. The program 'MEANS' was used to compute this average. The technique is very simple and straightforward as the ensemble mean was computed. This agrees with the

work done by the Air Force [Ref. 28] where they assumed that the samples are statistically independent, identically distributed Gaussian random variables. This is an over simplification as it is known that the vocal tract is slowly varying with the tract parameters changing only every 10 msec.

The short-term cepstral representation of the averaged waveform was computed using the program 'CEP'. In keeping with Cox and Robinson the waveform was segmented into 25 msec parts and each part was processed in sequence.

Finally the analytic signal representation of the waveform was computed using a FIR Hilbert transformer with 79 weights, and a lower cutoff frequency of .05. The frequency response of this filter is shown in Figure 3.3. This particular filter was chosen over the other two 79 weight filters because of its very small approximation error. The small approximation error does imply that the transition band of this filter is larger than the other two filters, however, this was deemed less important than the peak approximation error.

Examples of these three representations of the same utterances can be found in Figures 4.1 thru 4.30. These examples are of a male 30 years old, born and raised in eastern Pennsylvania, and a Naval cryptologic officer. In order to display all 7680 points on one graph the waveform

was first normalized, then divided into four 1920 point parts. Each part was biased by (N-1) * 2, where N = 1, 2, 3, 4, to permit graphing by the four segments on one page. The graphs should be read from left to right, top to bottom.

C. DECISION ALGORITHM

Once the speech had been processed a decision algorithm had to be formulated to classify utterances based on the patterns collected. All of the isolated word recognizers use a form of classical pattern recognition to classify utterances. The VRM system uses a nearest neighbor algorithm with a variable threshold. If no utterance is within the distance specified by the threshold, an unable to classify message is issued.

The nearest neighbor rule is an example of the pooled form of the nearest neighbor rule [Ref. 29]. For the two class case, a hemisphere is formed around the vector $\underline{y}$ to include k total samples regardless of their class. Thus $k_1 + k_2 = k$, where k_i equals the number of vectors belonging to class i. The quotient k_1/k_2 is formed and compared to one. If $k_1/k_2 > 1$, then this implies there are more class one vectors in the hemisphere around $\underline{y}$ and the vector $\underline{y}$ is said to belong to class one. If the converse of the inequality is true, $k_1/k_2 < 1$, then $\underline{y}$ is said to belong to class two. The probability of error for the case k=1 is less than twice the minimum probability of error for any decision rule.

The nearest neighbor rule was employed to classify the utterances. Using the program 'DEC', the Euclidean distance between a test vector and the stored patterns was computed. The results of this pattern matching are discussed in the next chapter.

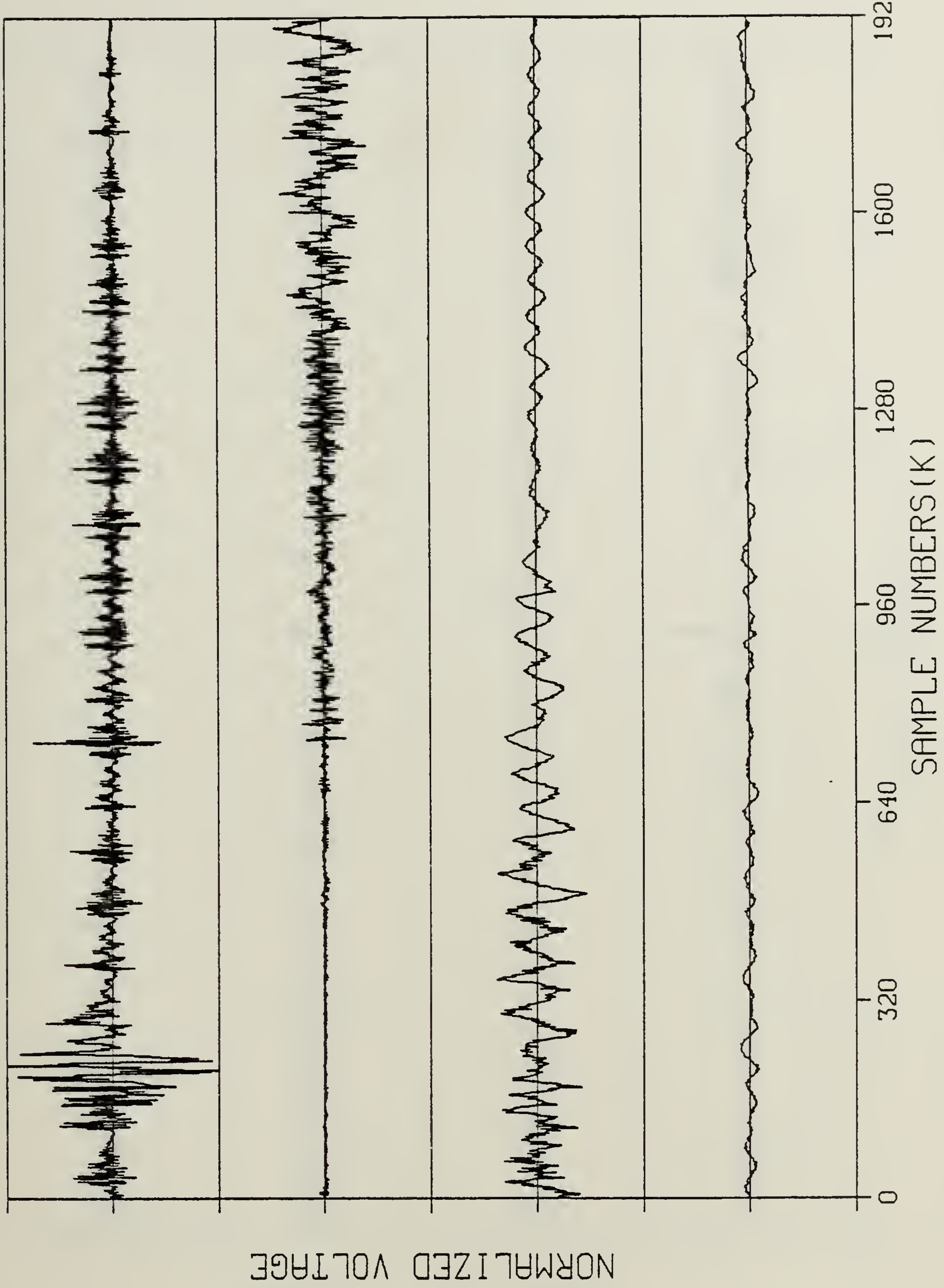

Figure 4.1. Sampled Waveform, Zero

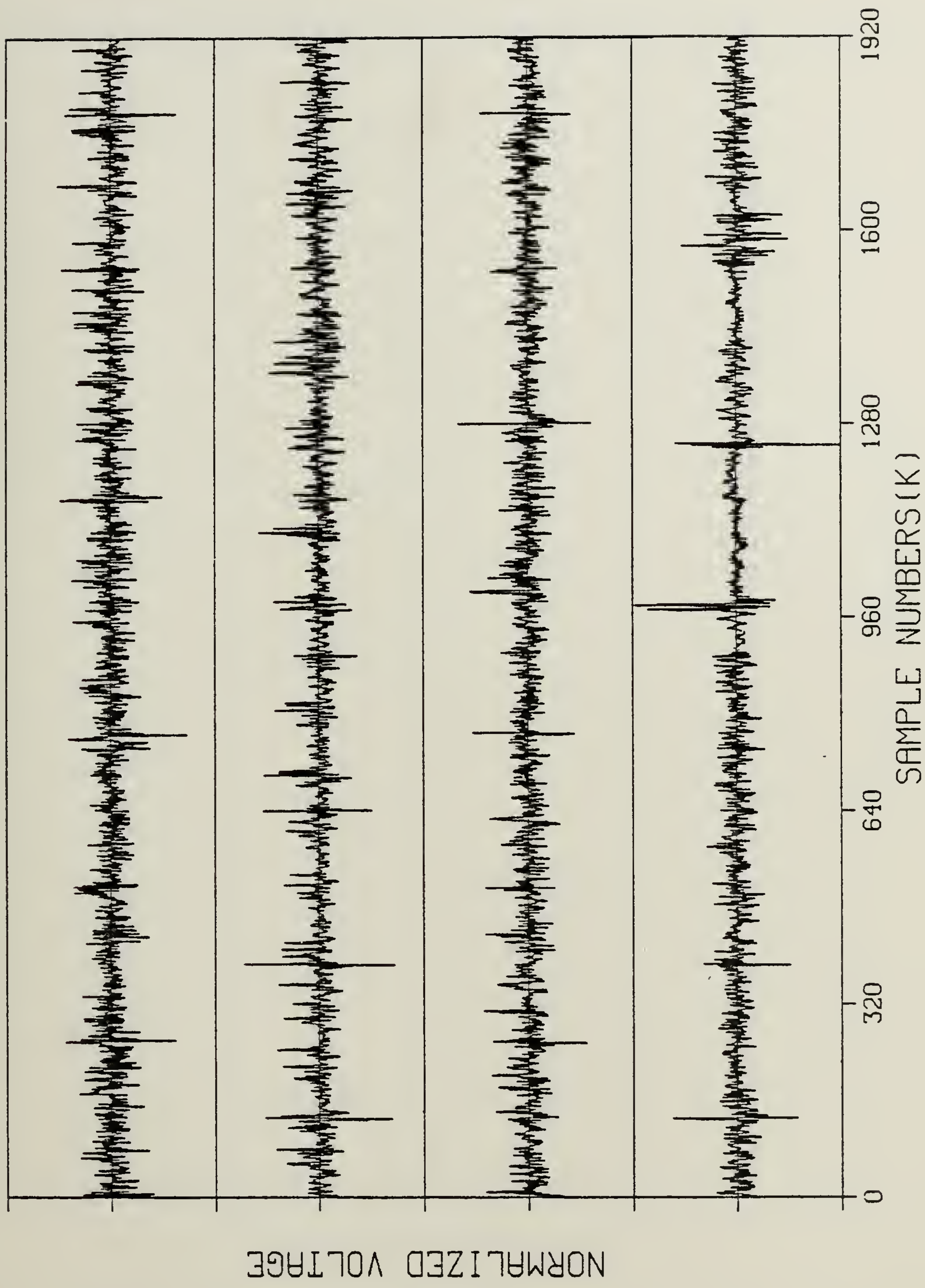

Figure 4.2. Analytic Representation of Zero

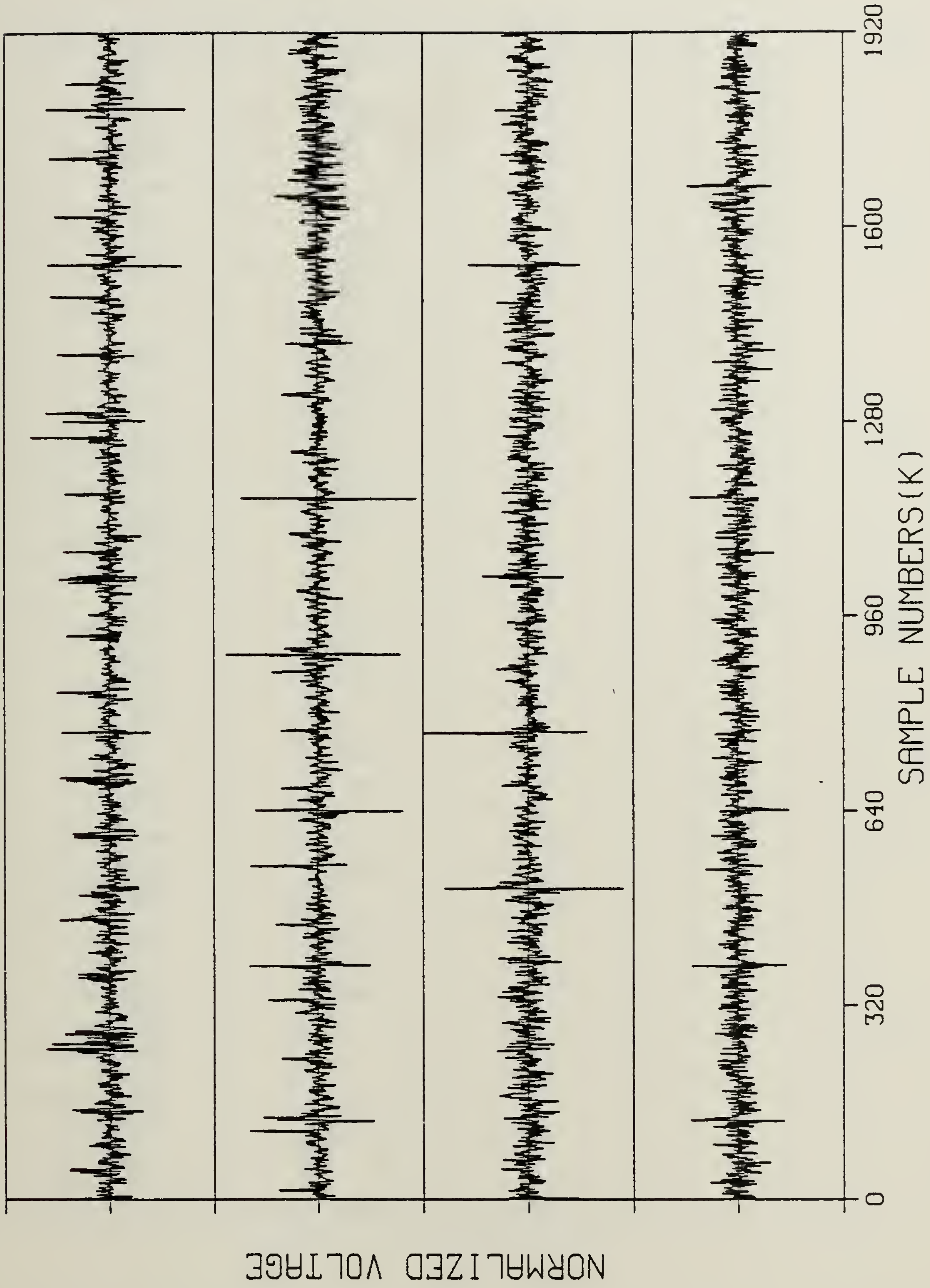

Figure 4.3. Cepstral Representation of Zero

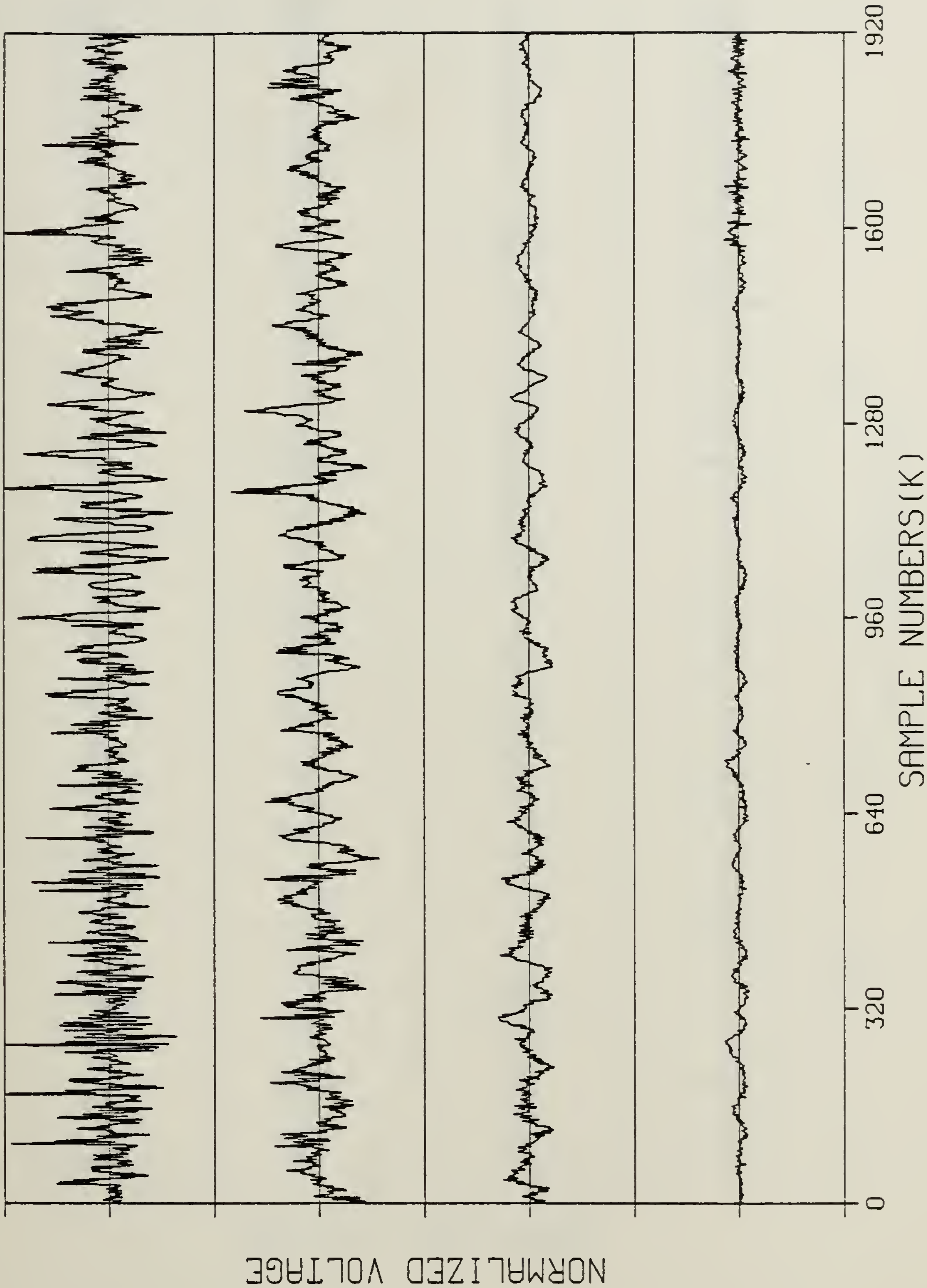

Figure 4.4. Sampled Waveform, One

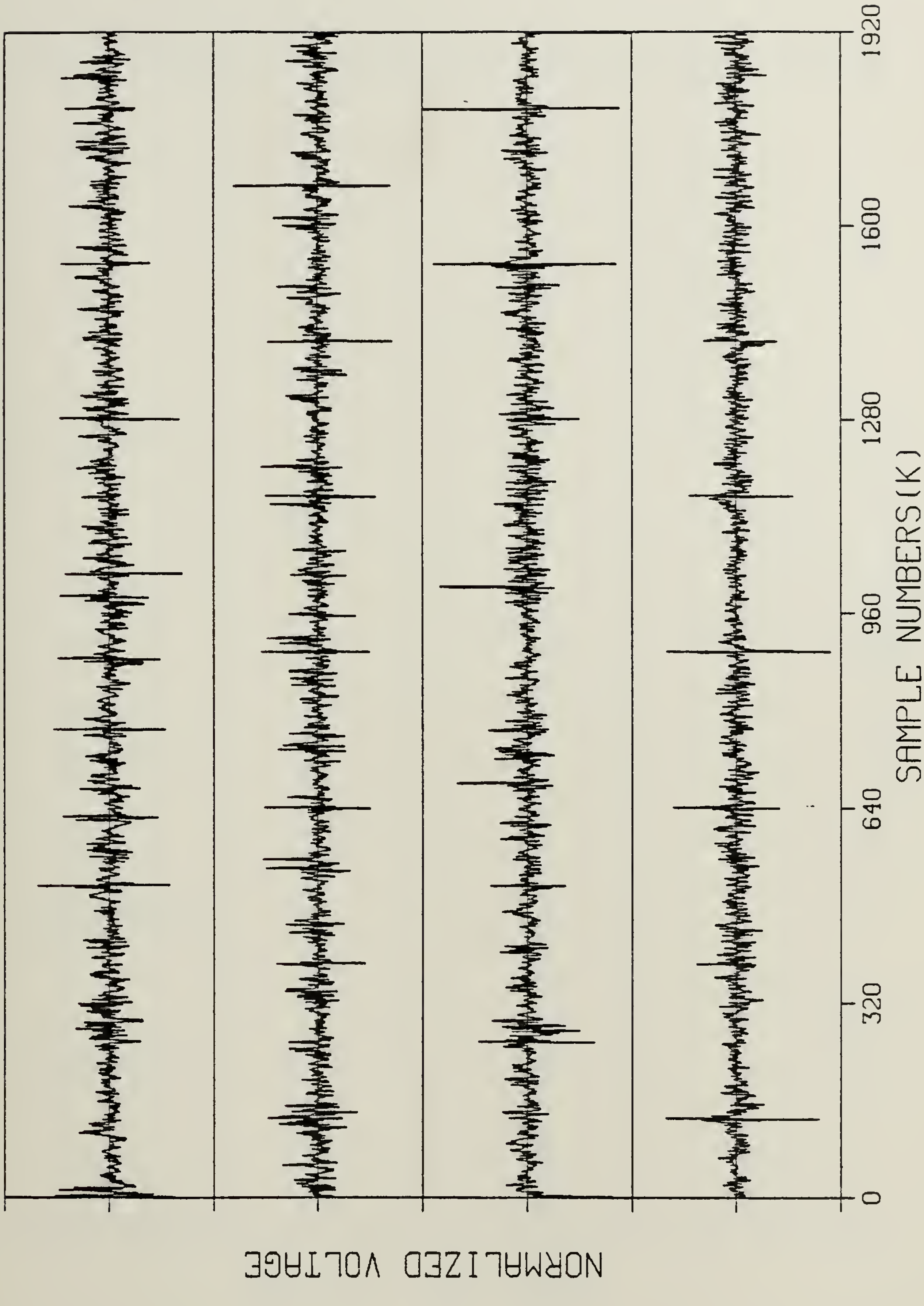

Figure 4.5. Analytic Representation of One

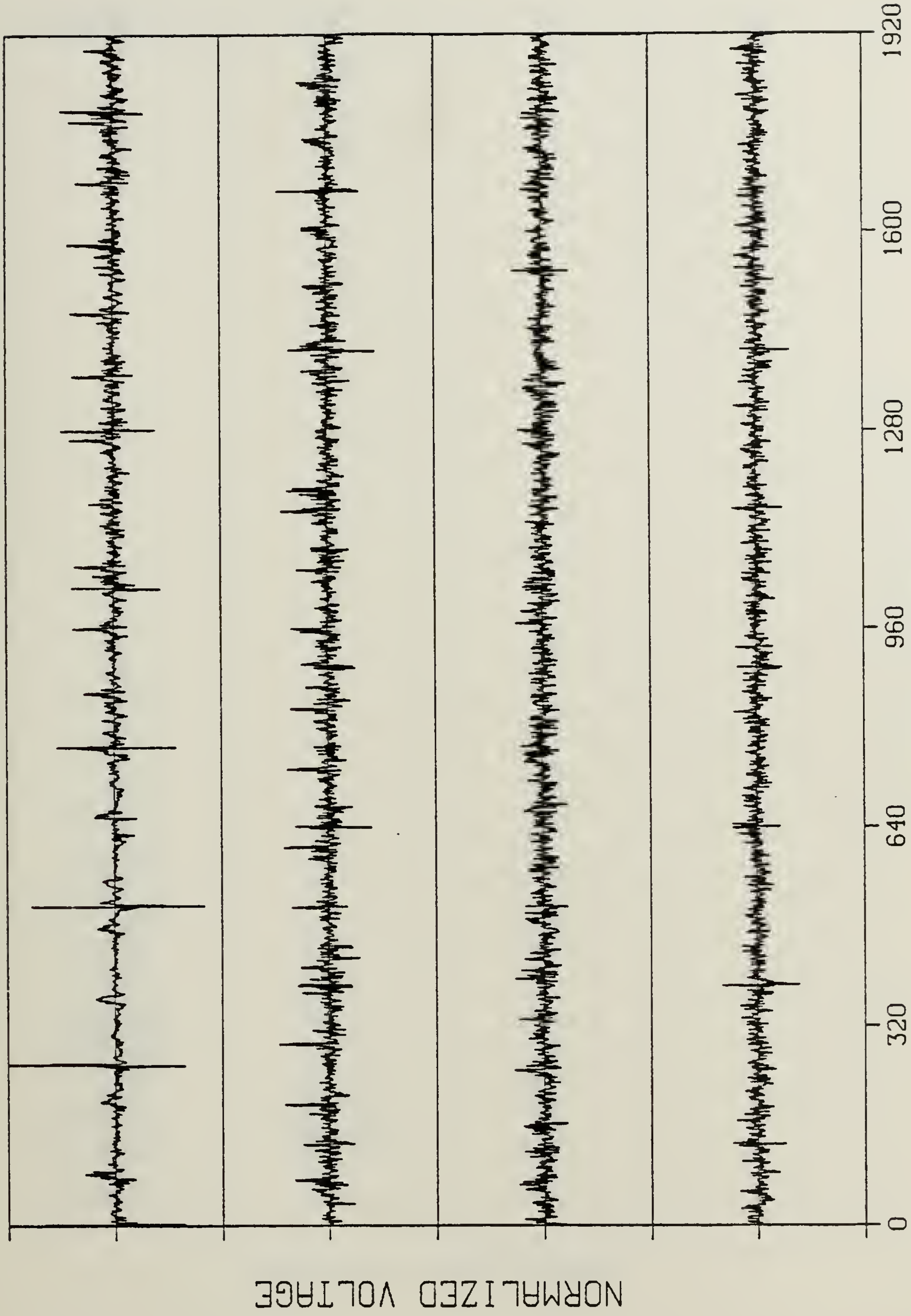

Figure 4.6. Cepstral Representation of One

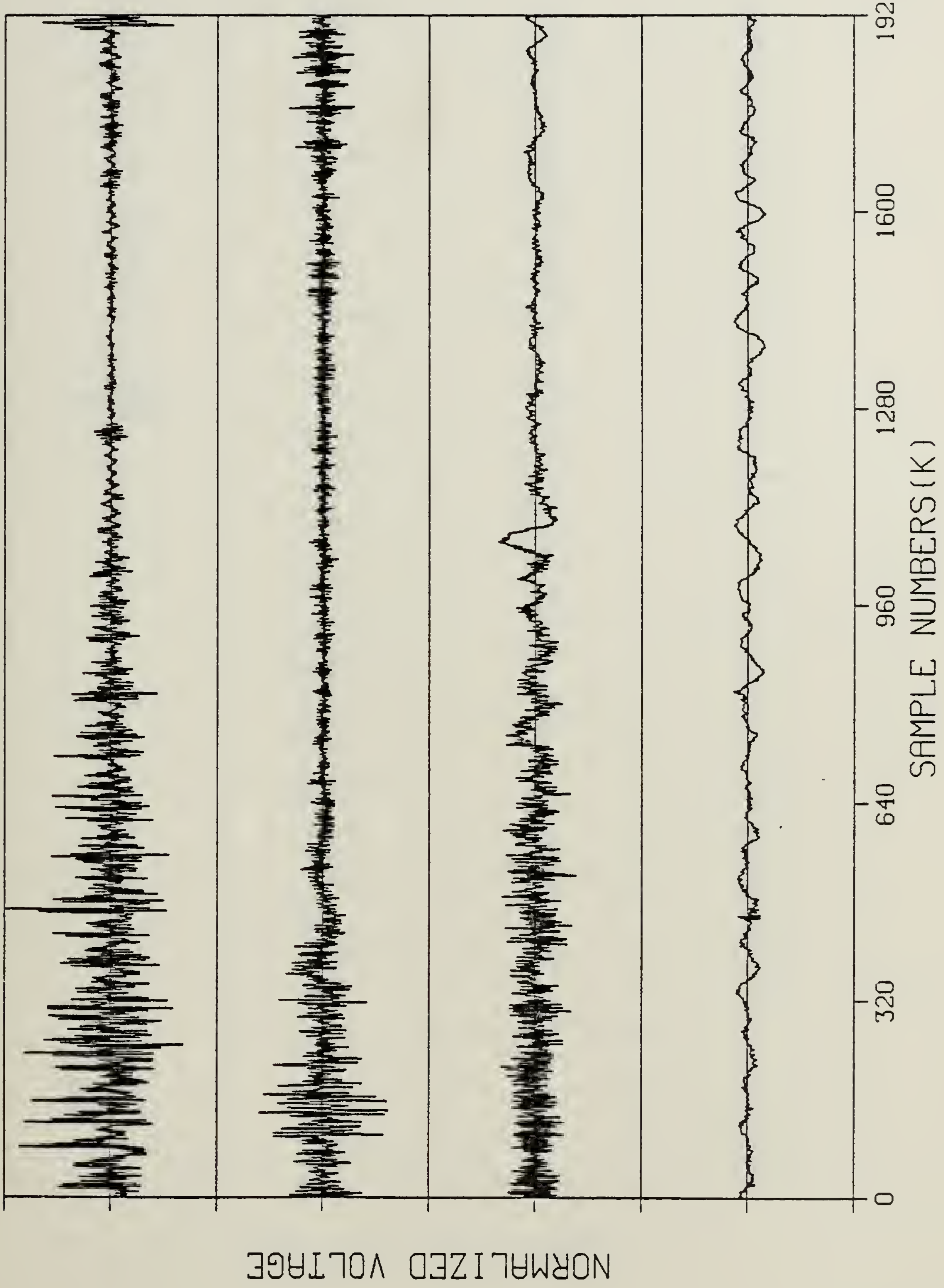

Figure 4.7. Sampled Waveform, Two

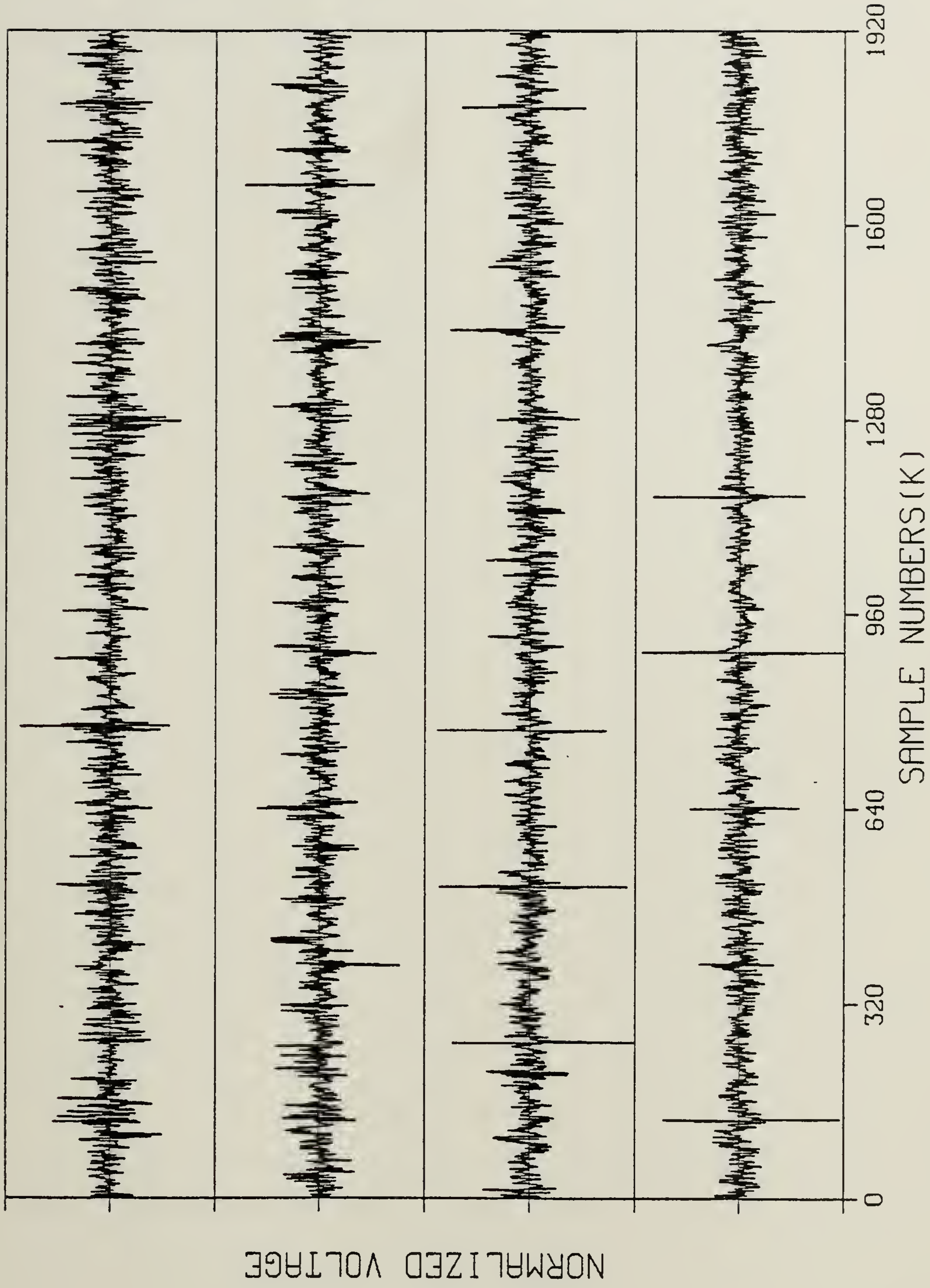

Figure 4.8. Analytic Representation of Two

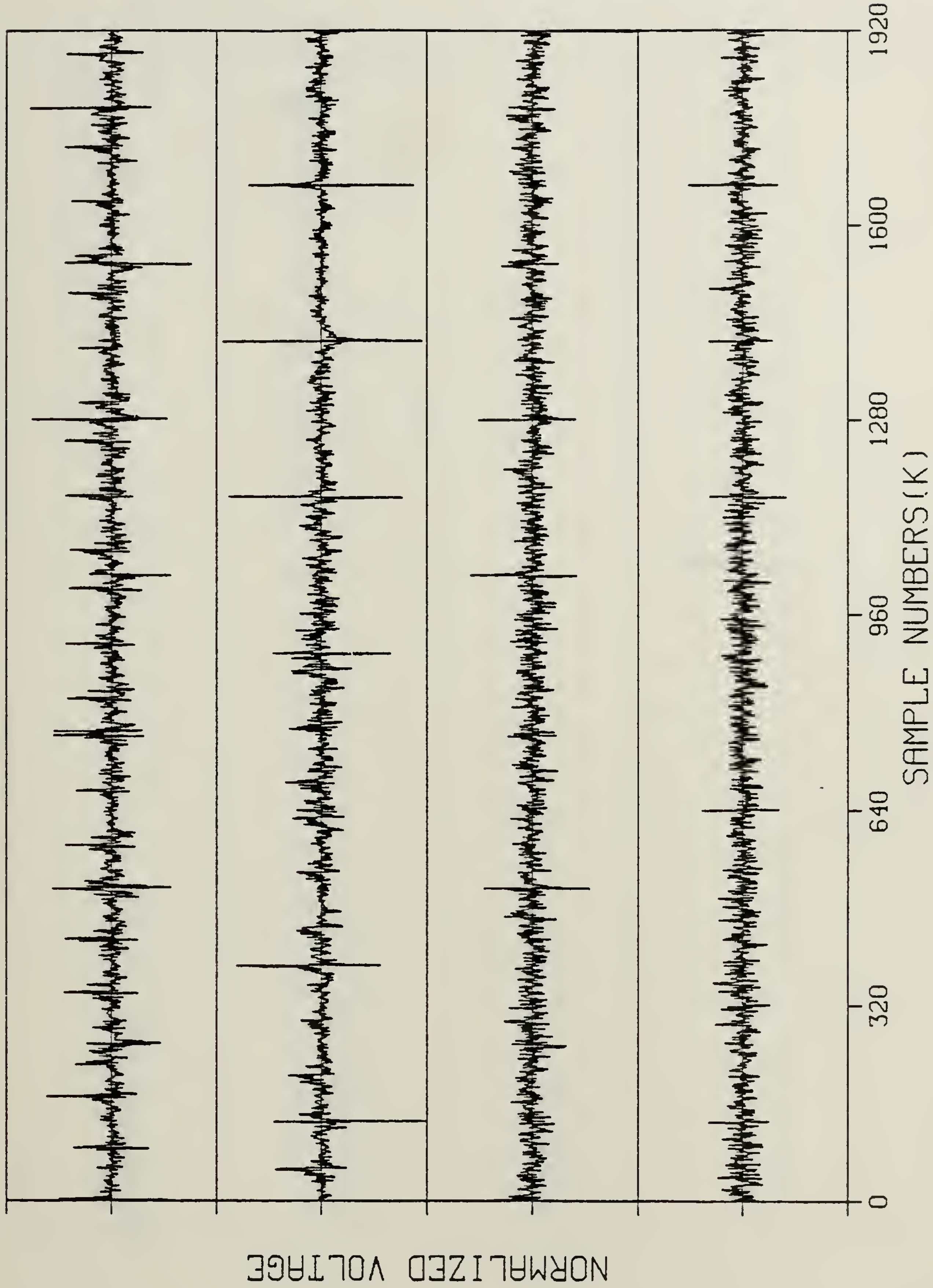

Figure 4.9. Cepstral Representation of Two

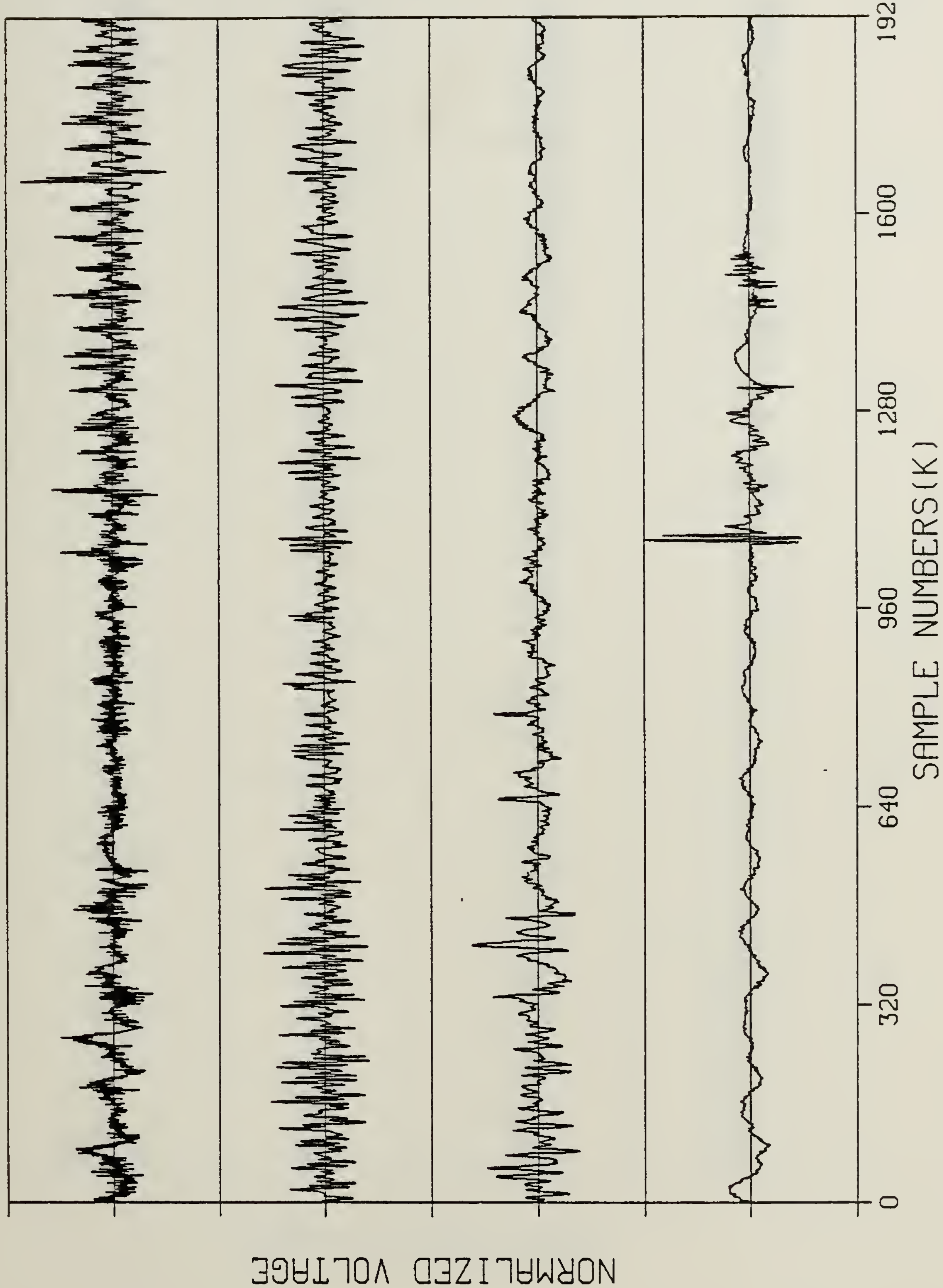

Figure 4.10. Sampled Waveform, Three

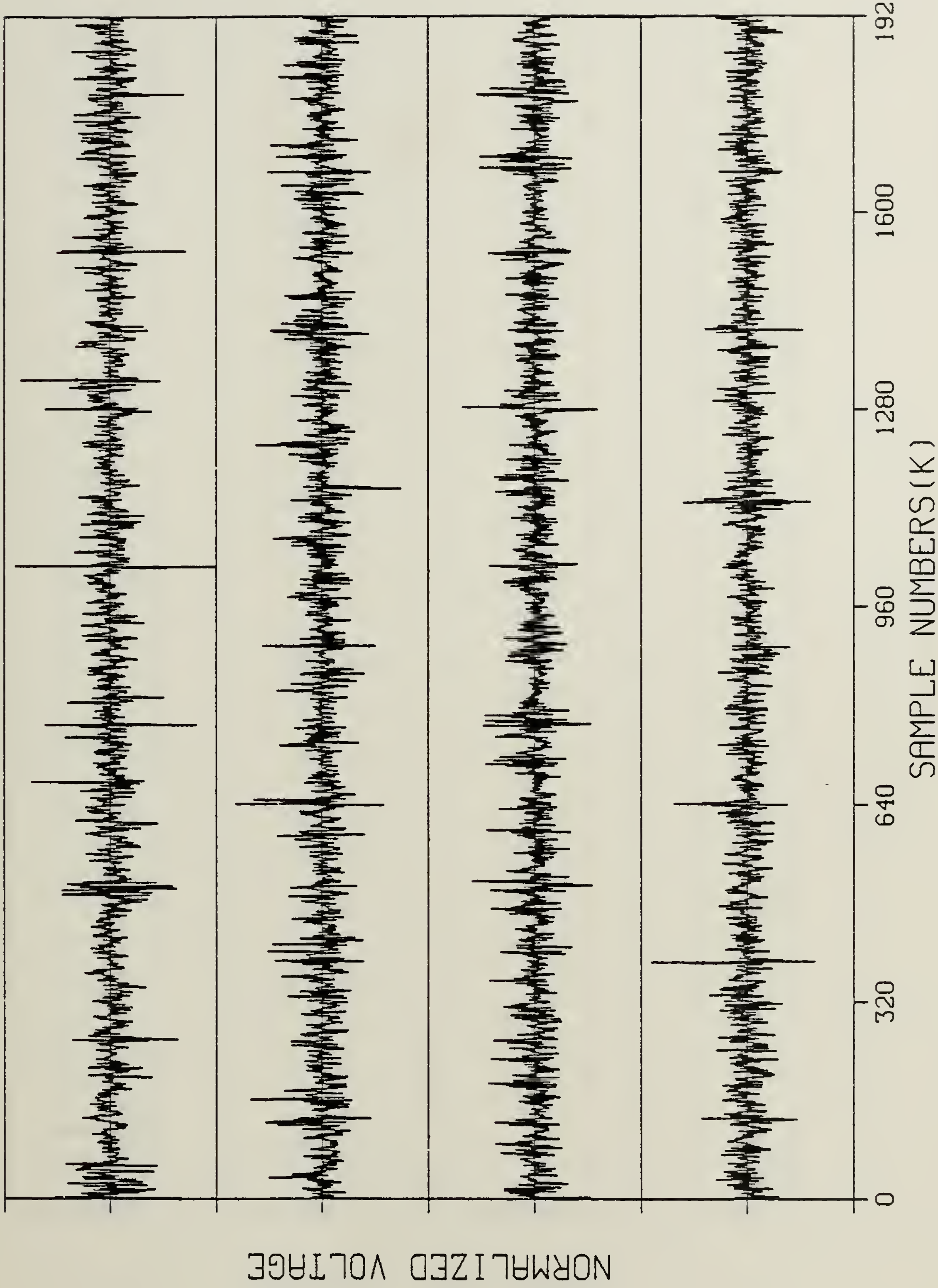

Figure 4.11. Analytic Representation of Three

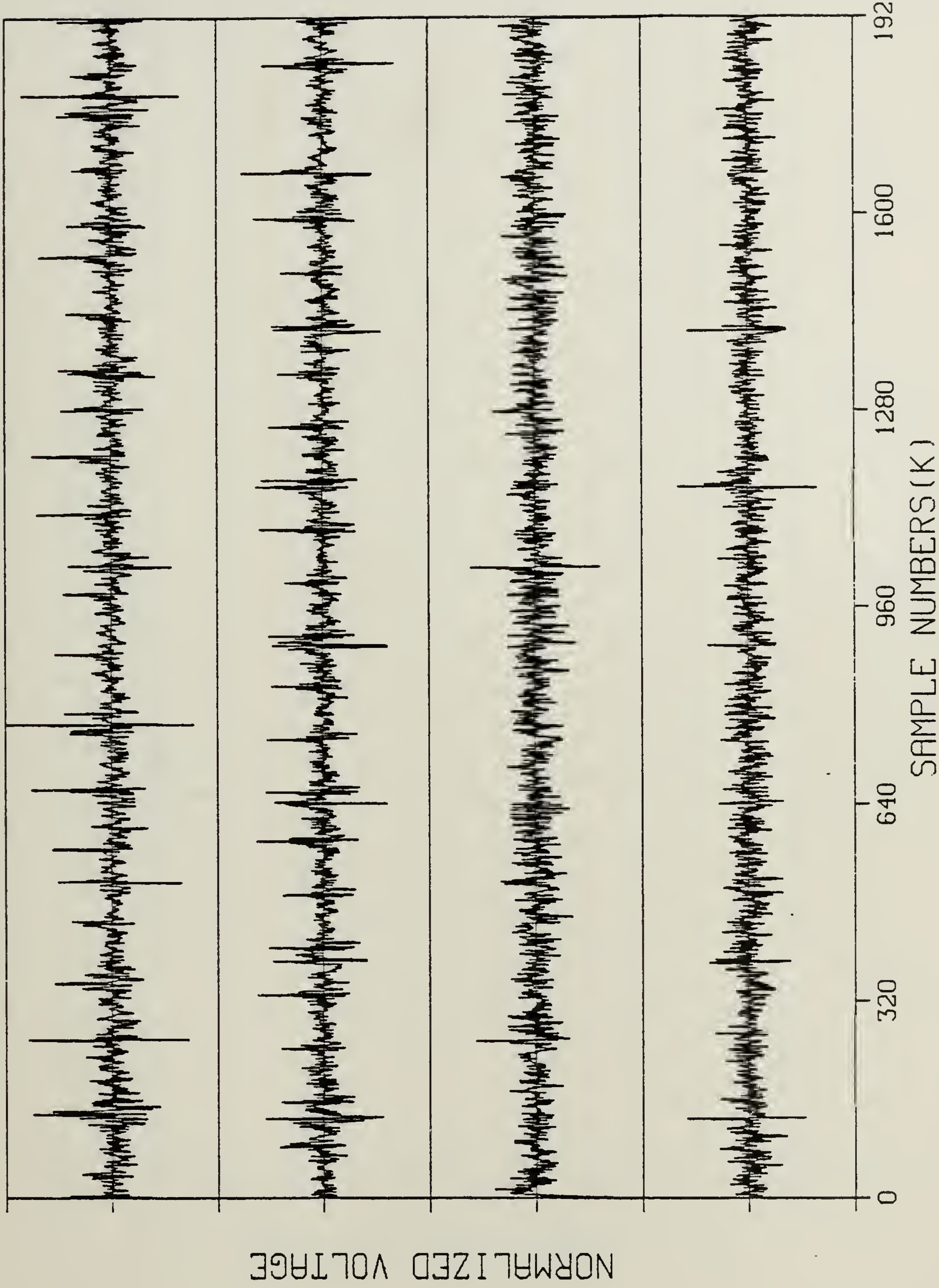

Figure 4.12. Cepstral Representation of Three

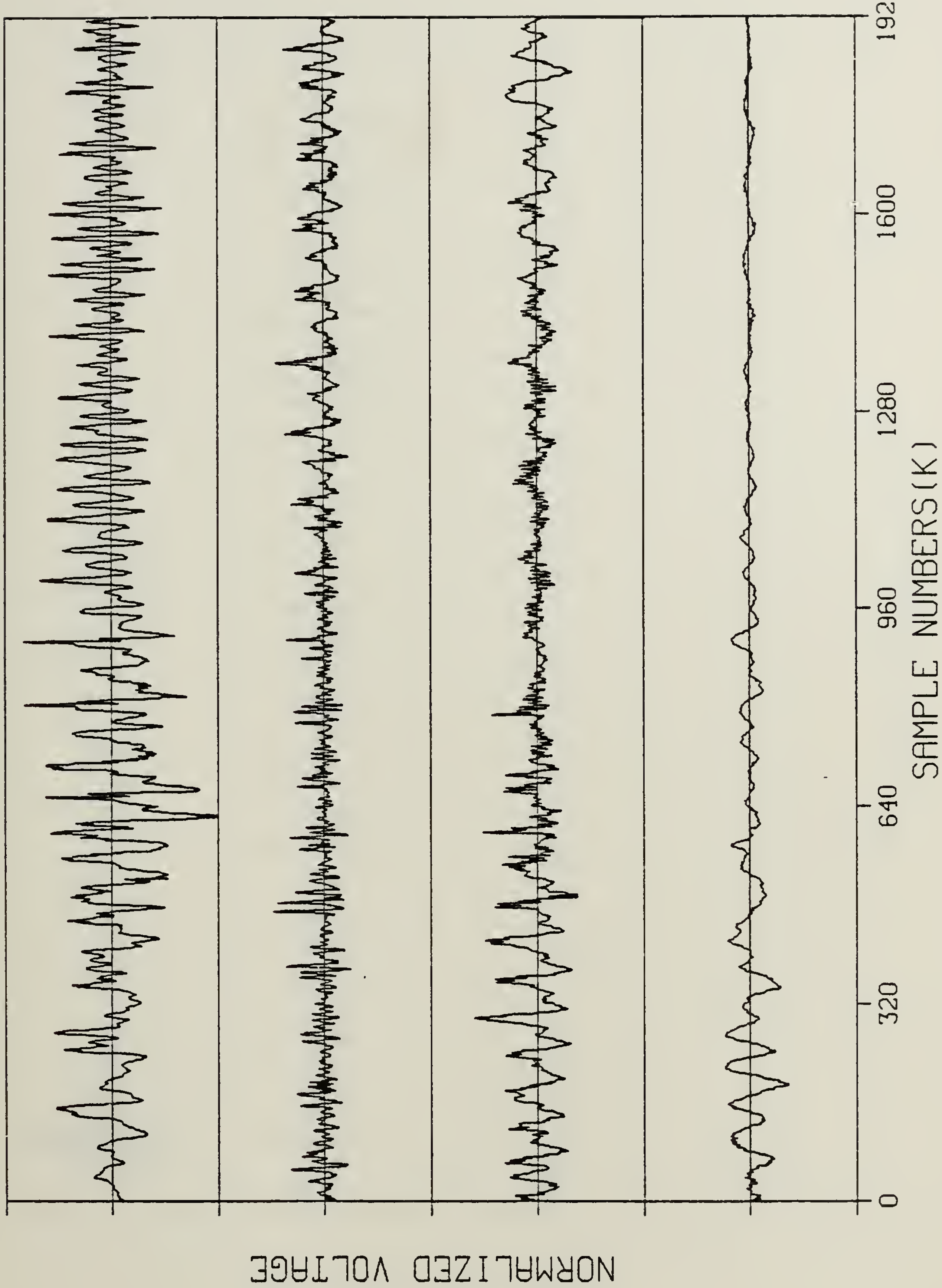

Figure 4.13. Sampled Waveform, Four

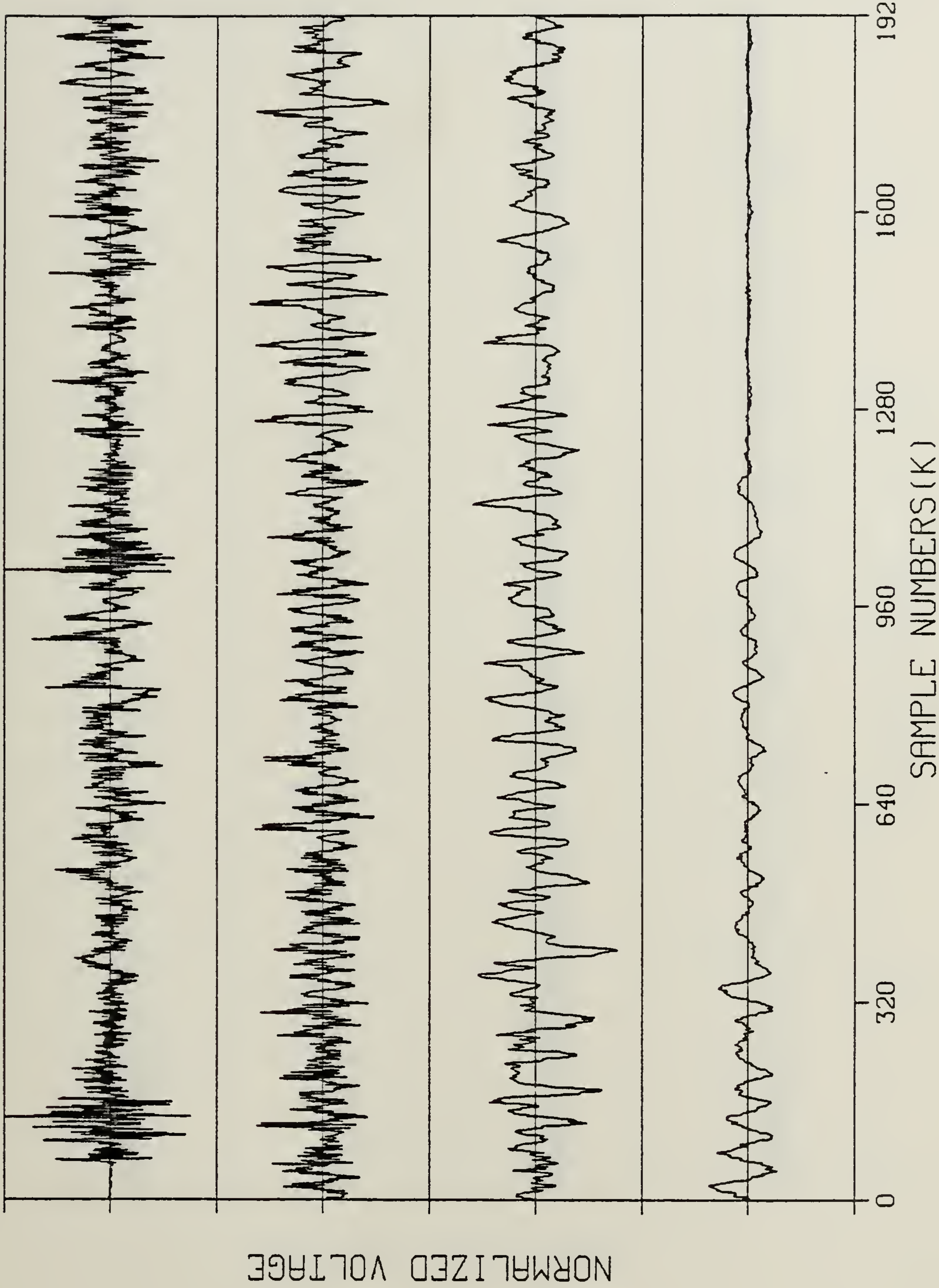

Figure 4.14. Analytic Representation of Four

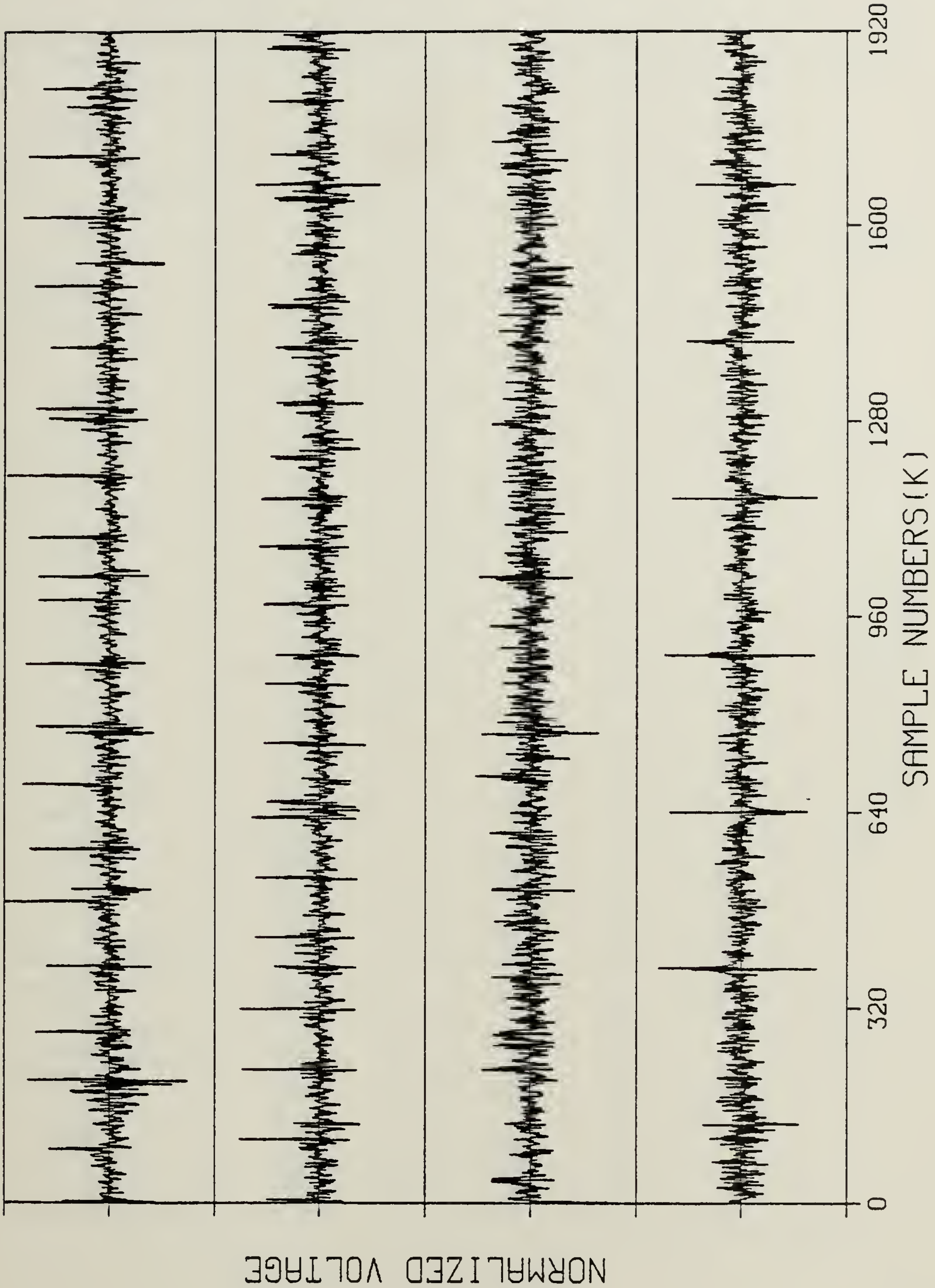

Figure 4.15. Cepstral Representation of Four

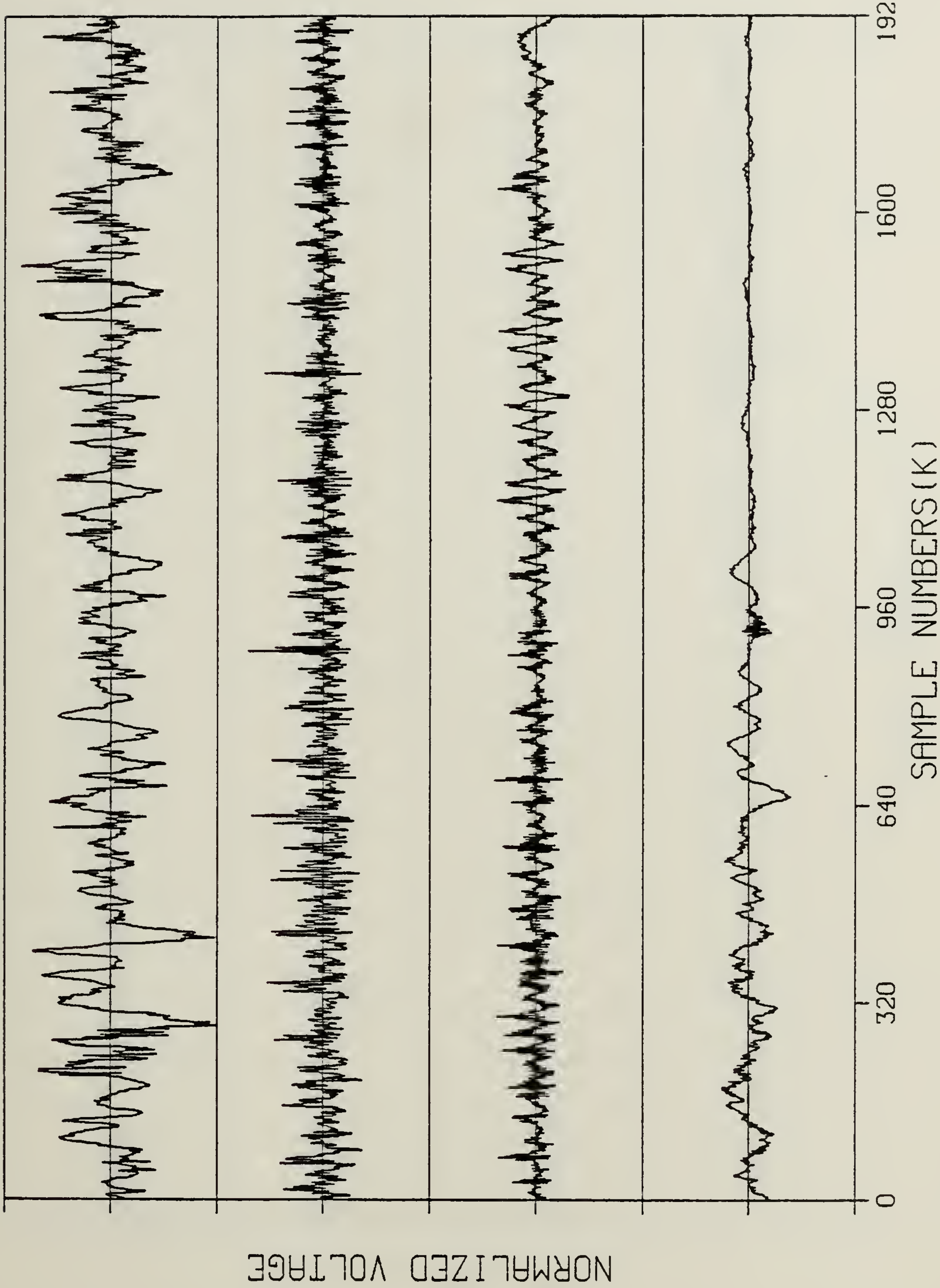

Figure 4.16. Sampled Waveform, Five

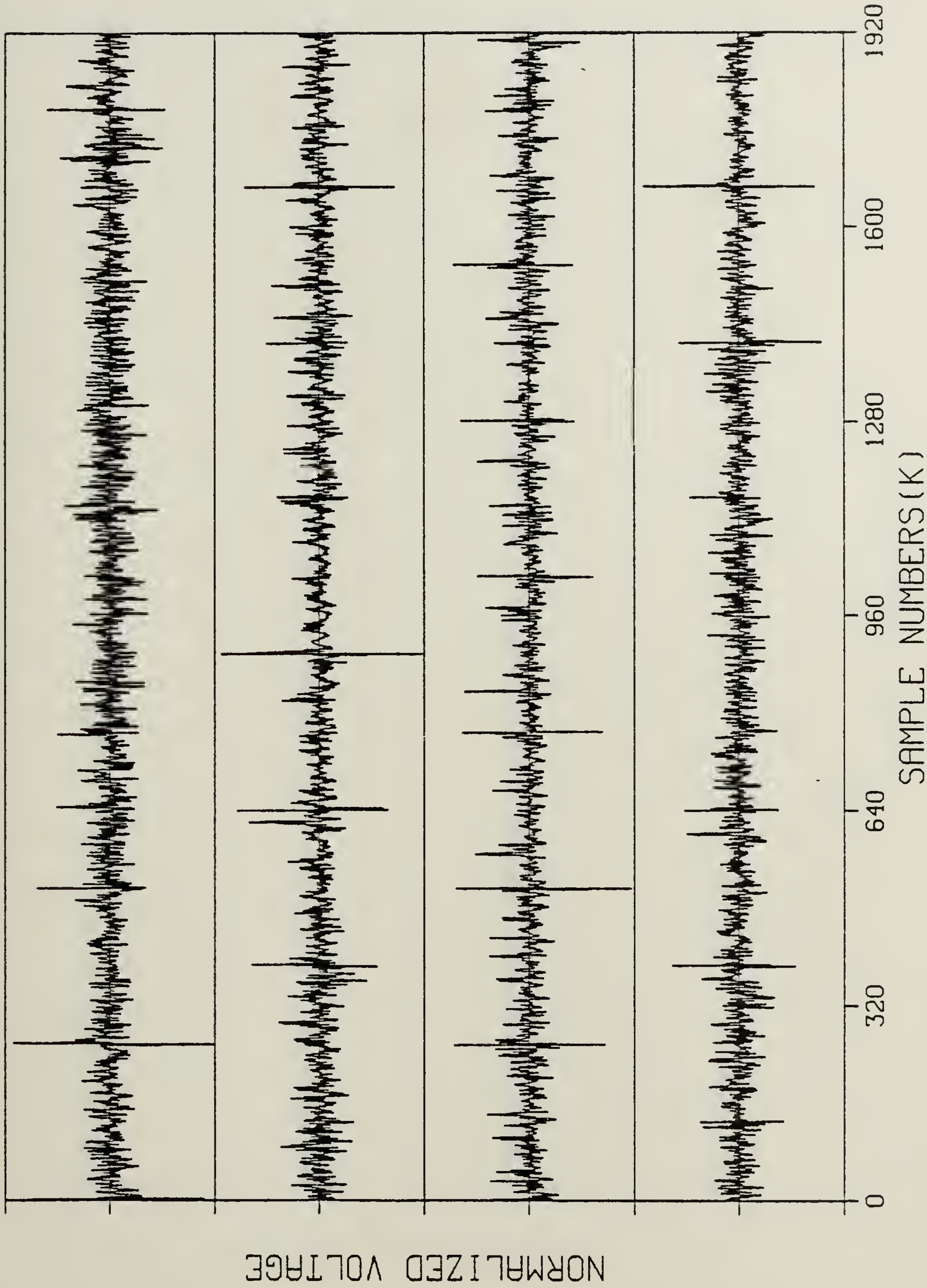

Figure 4.17. Analytic Representation of Five

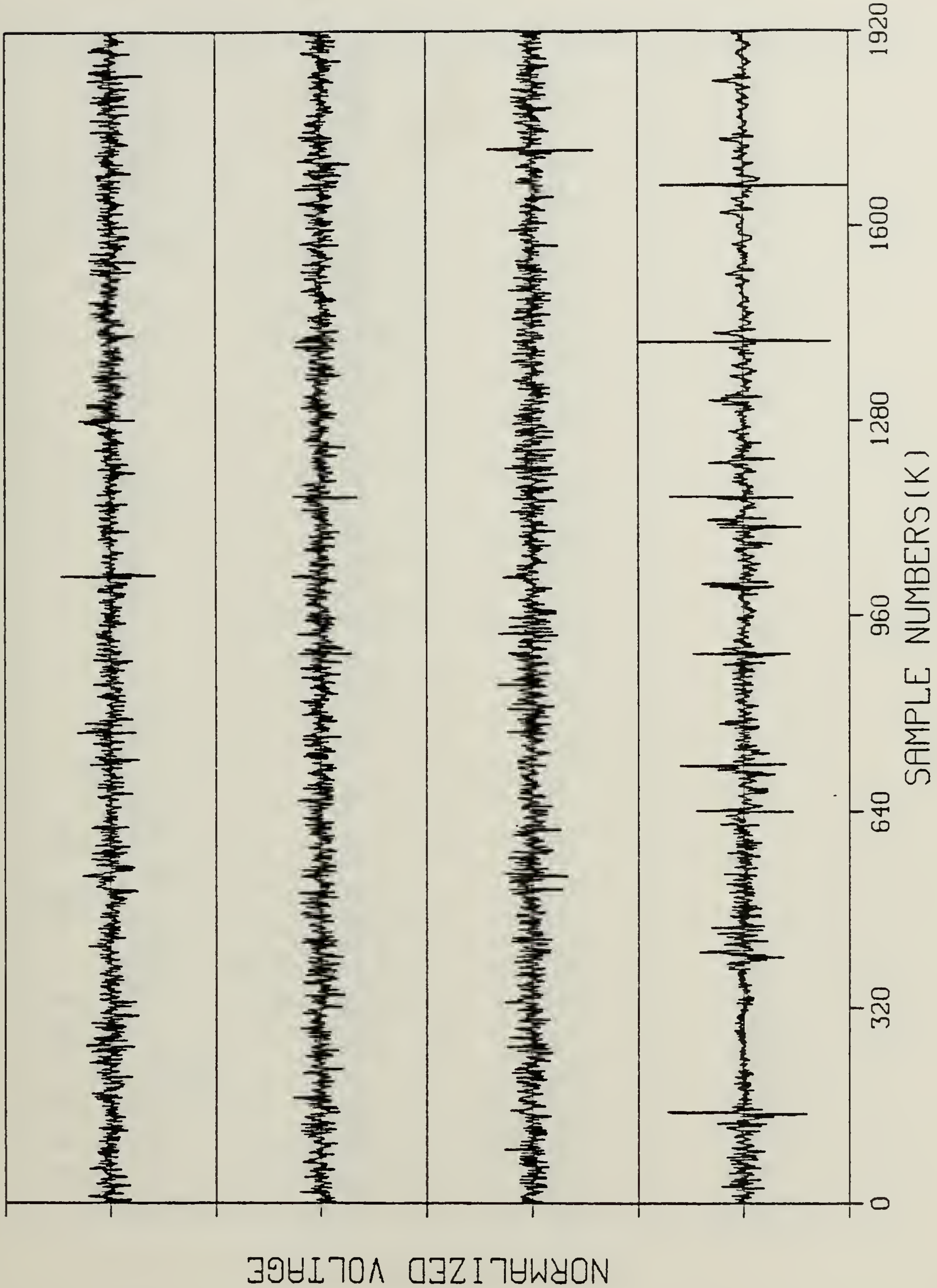

Figure 4.18. Cepstral Representation of Five

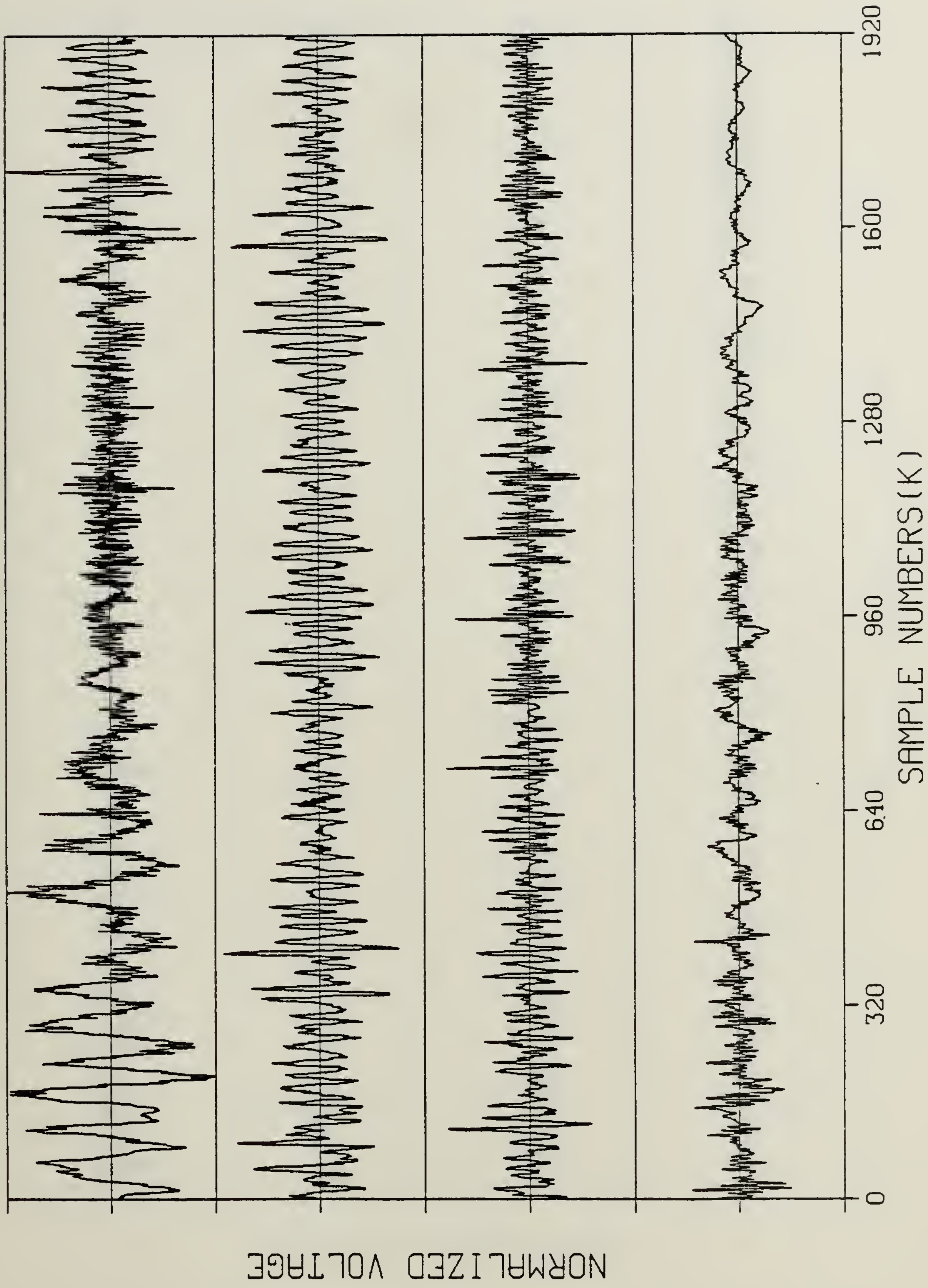

Figure 4.19. Sampled Waveform, Six

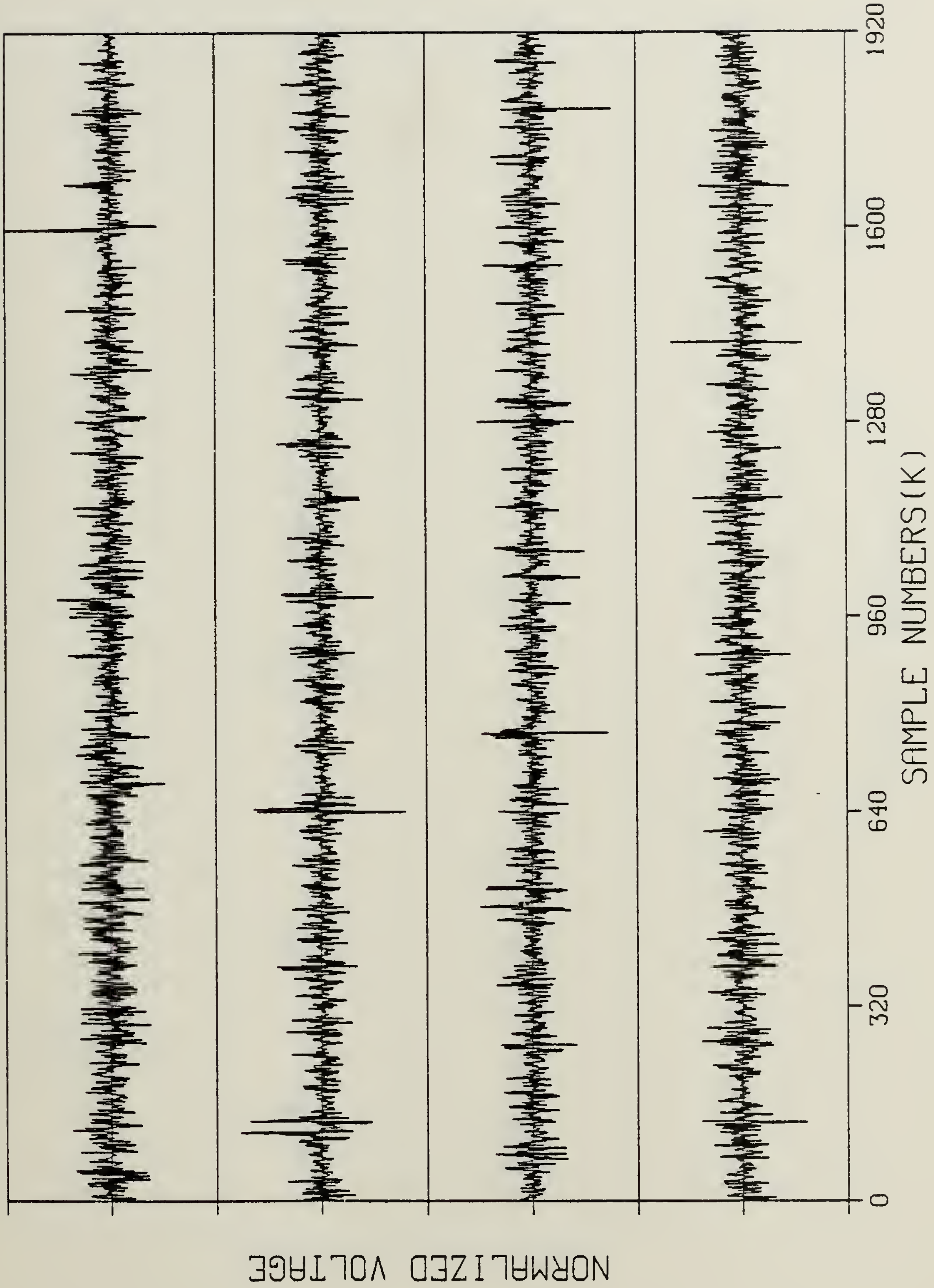

Figure 4.20. Analytic Representation of Six

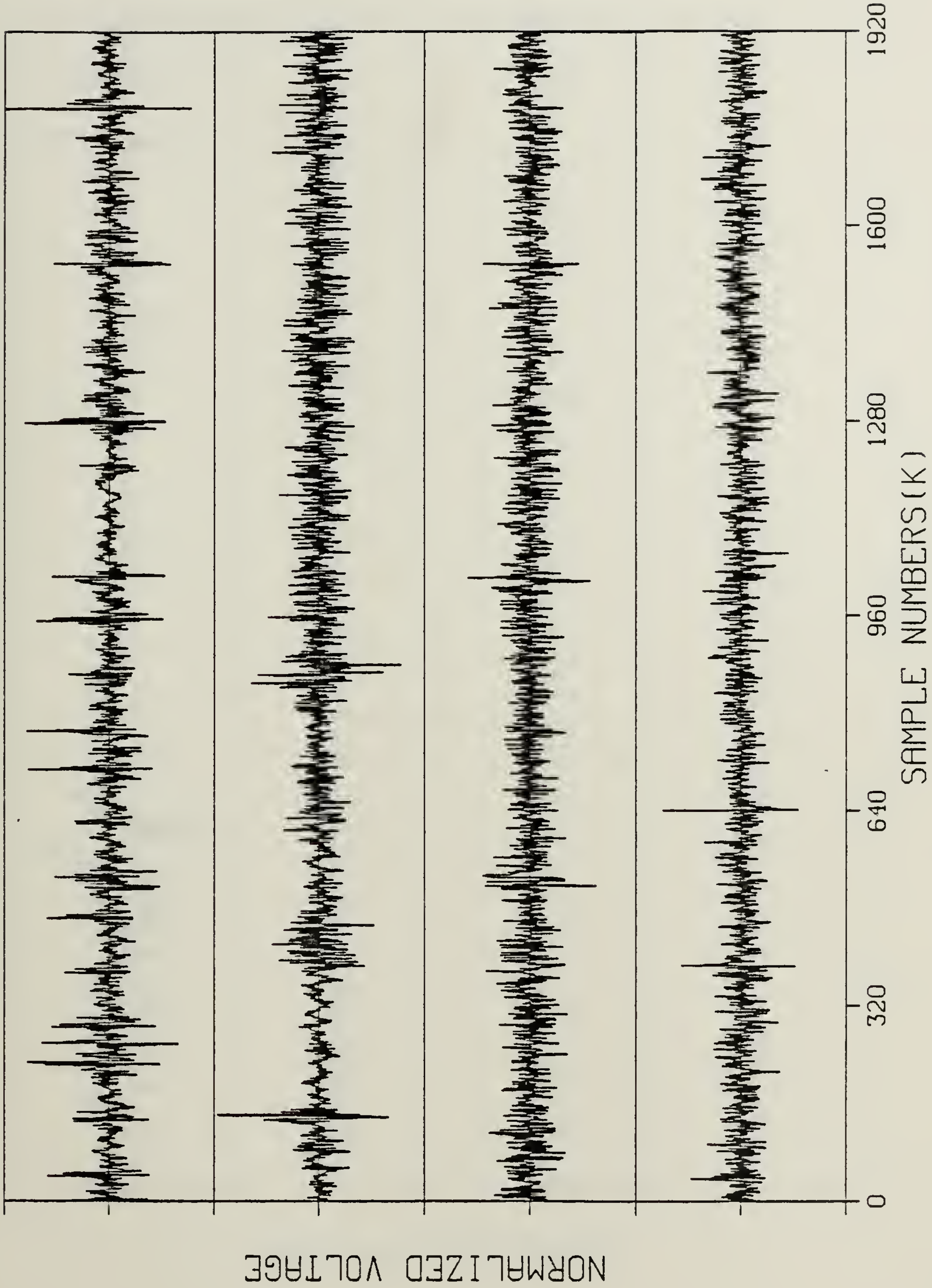

Figure 4.21. Cepstral Representation of Six

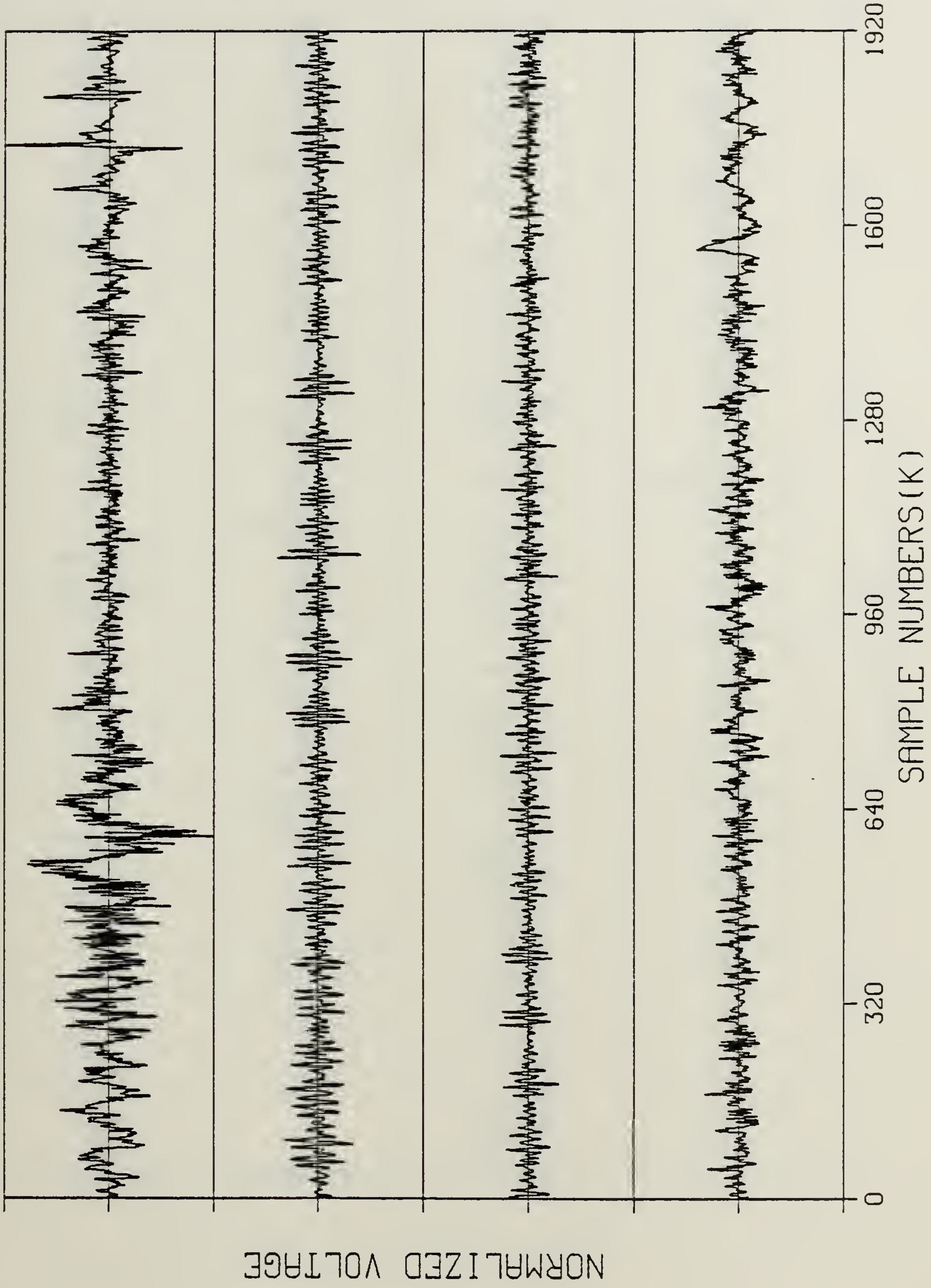

Figure 4.22. Sampled Waveform, Seven

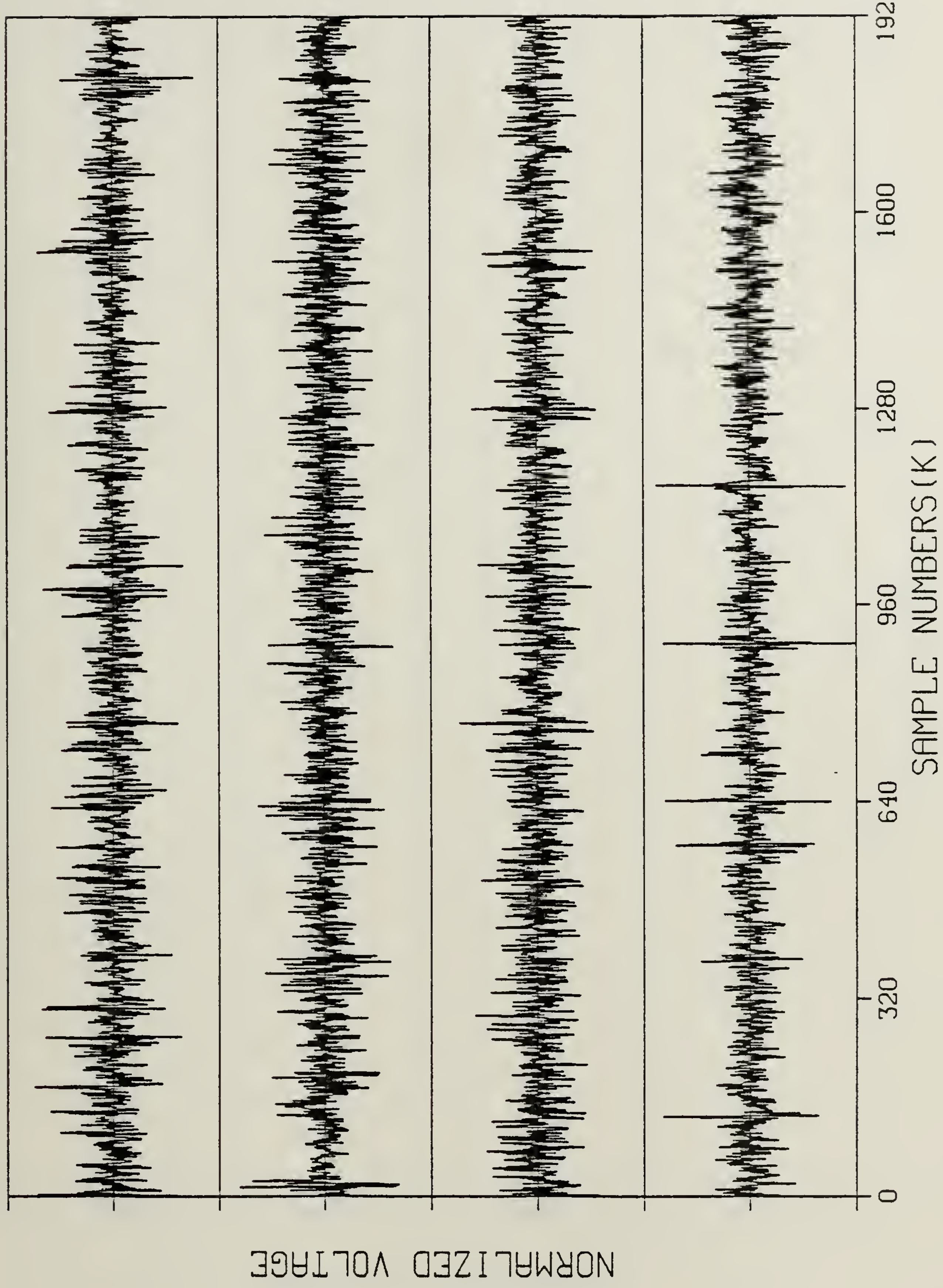

Figure 4.23. Analytic Representation of Seven

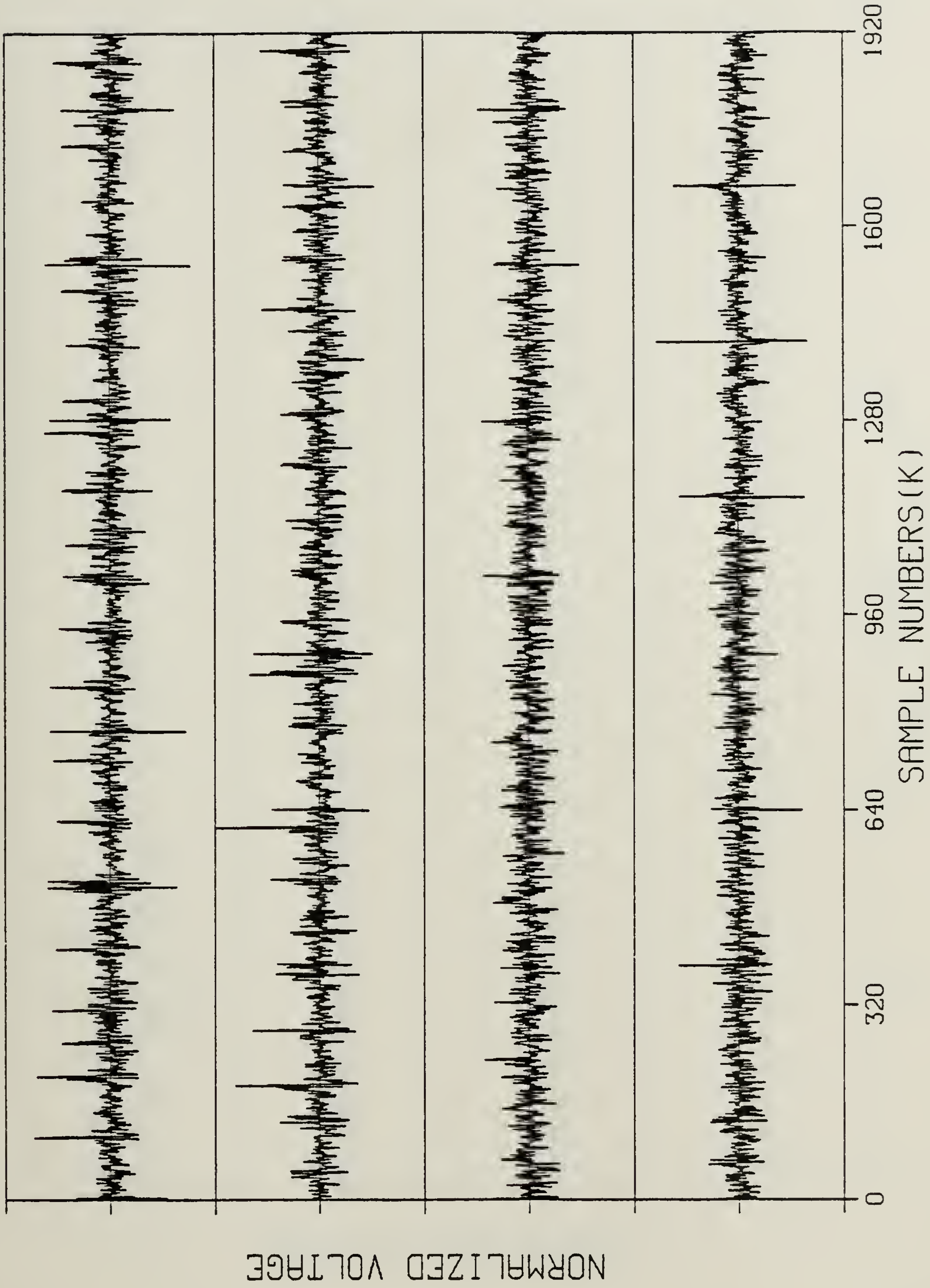

Figure 4.24. Cepstral Representation of Seven

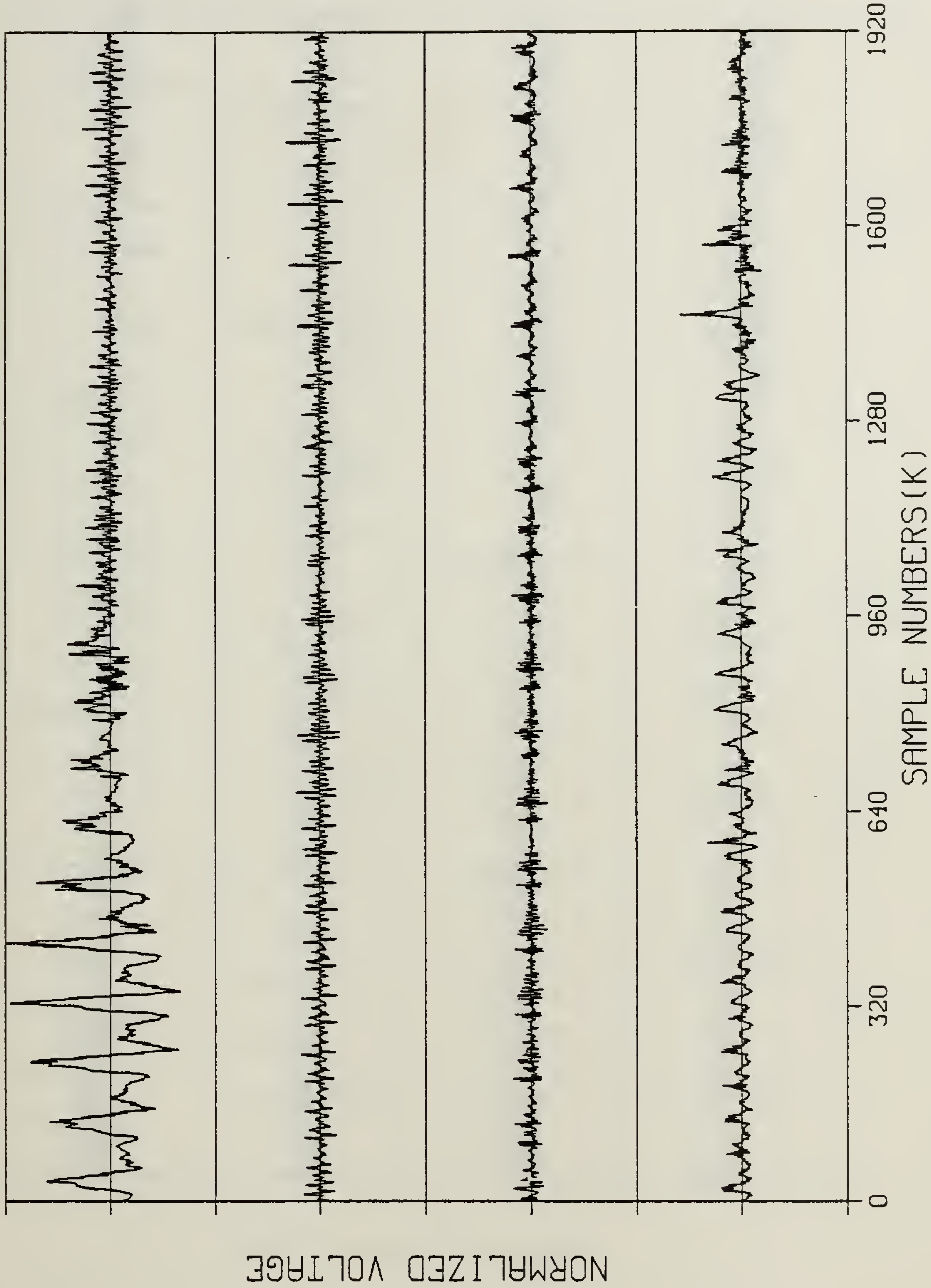

Figure 4.25. Sampled Waveform, Eight

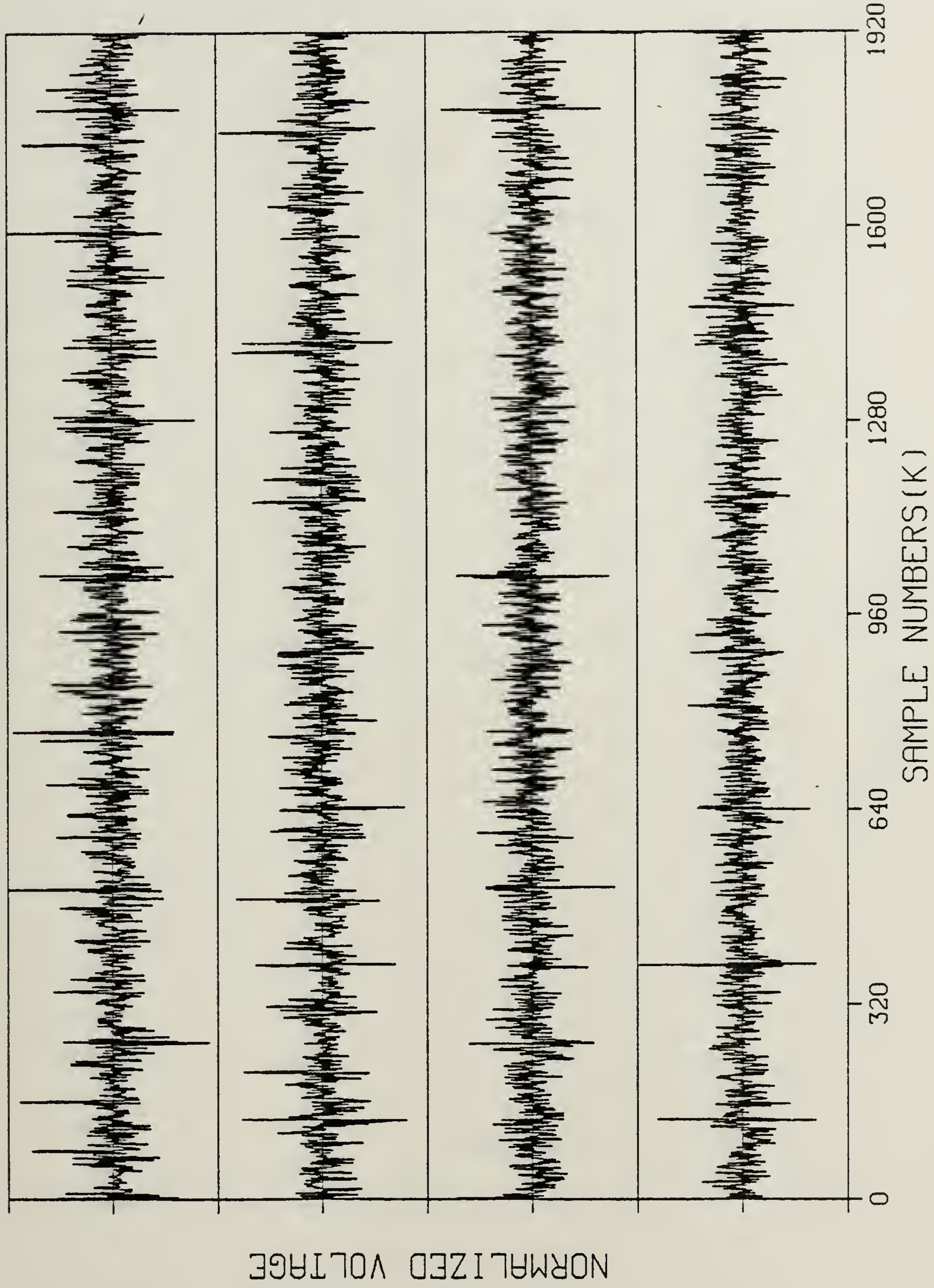

Figure 4.26. Analytic Representation of Eight

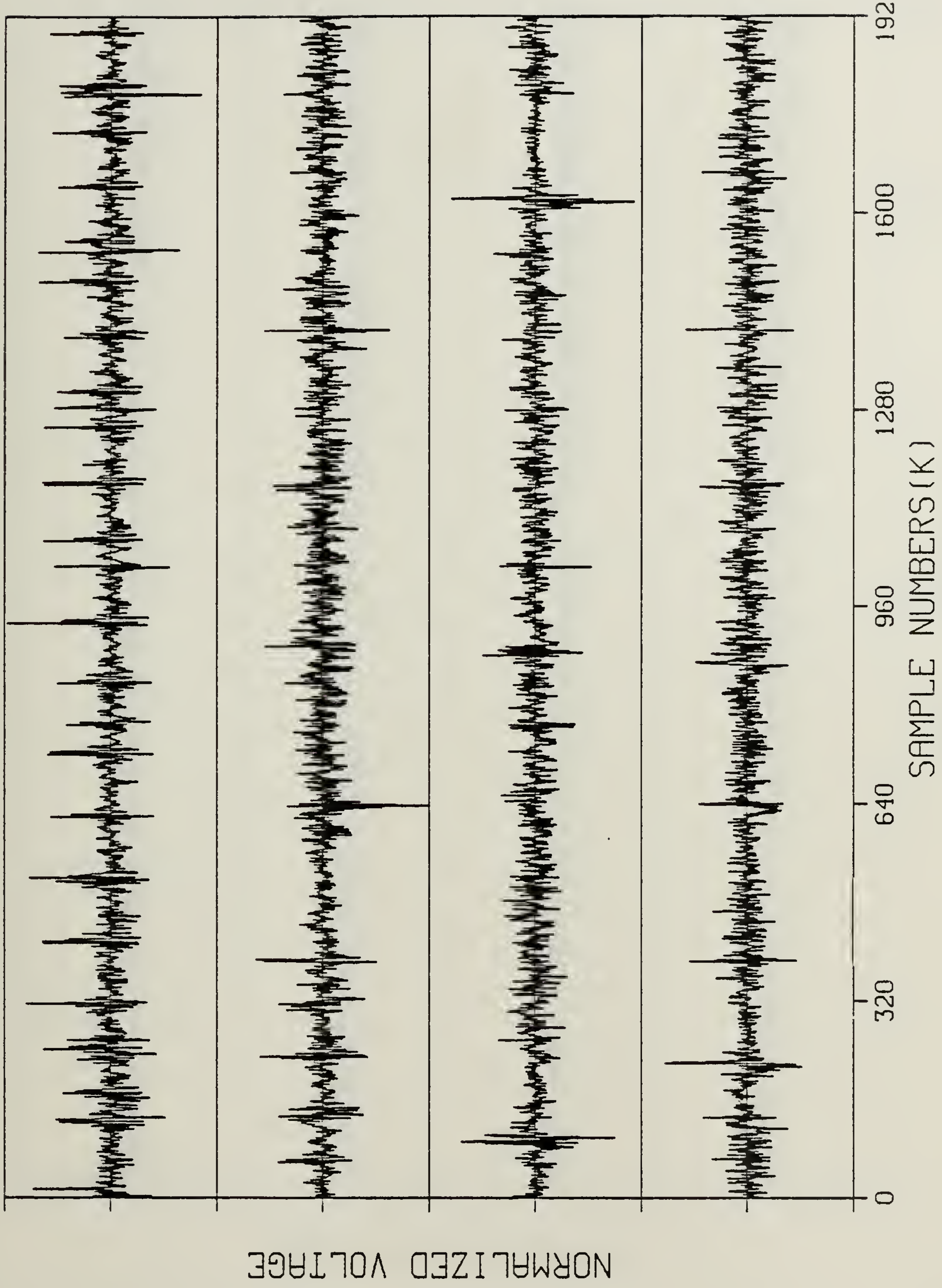

Figure 4.27. Cepstral Representation of Eight

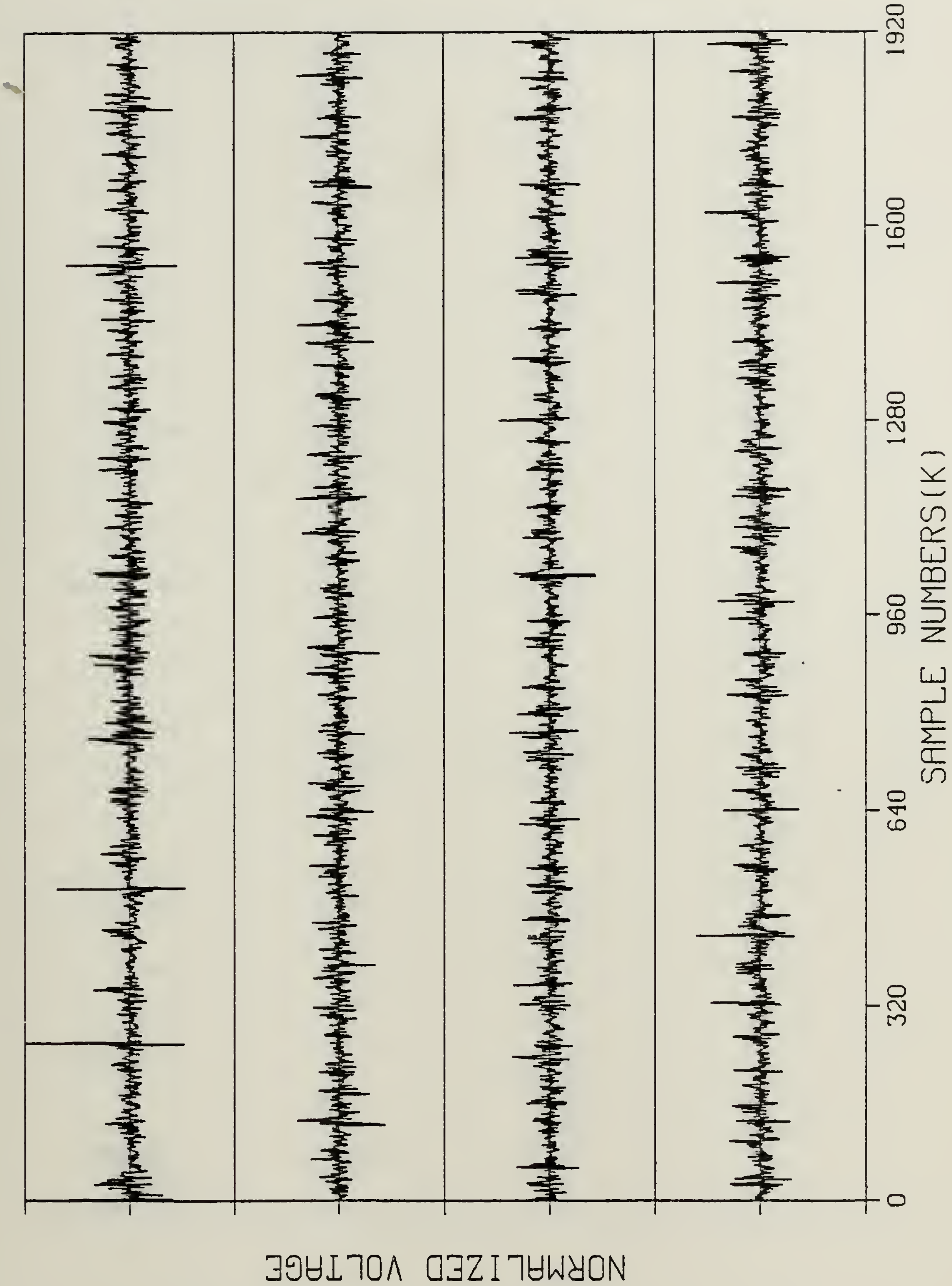

Figure 4.28. Sampled Waveform, Nine

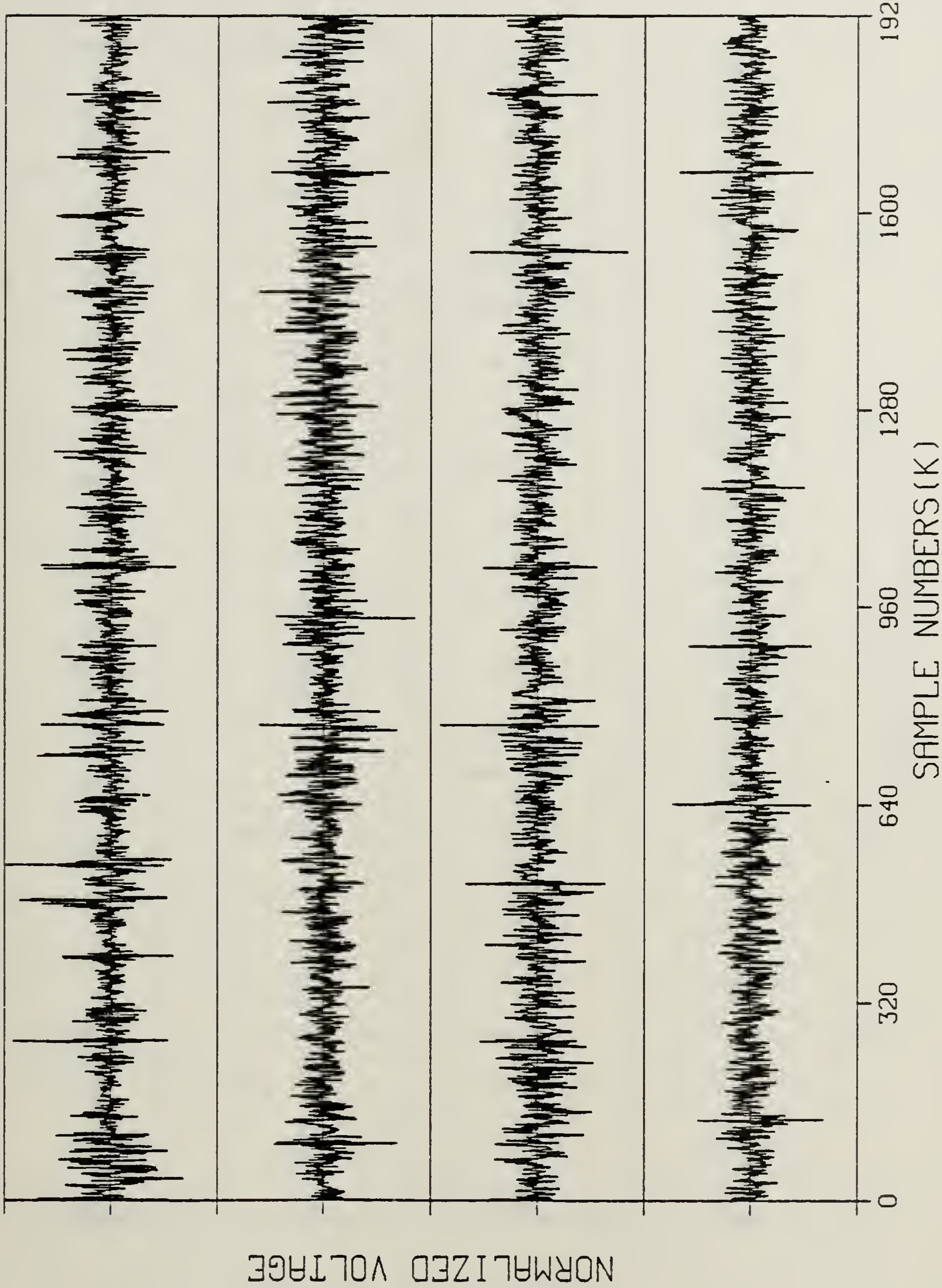

Figure 4.29. Analytic Representation of Nine

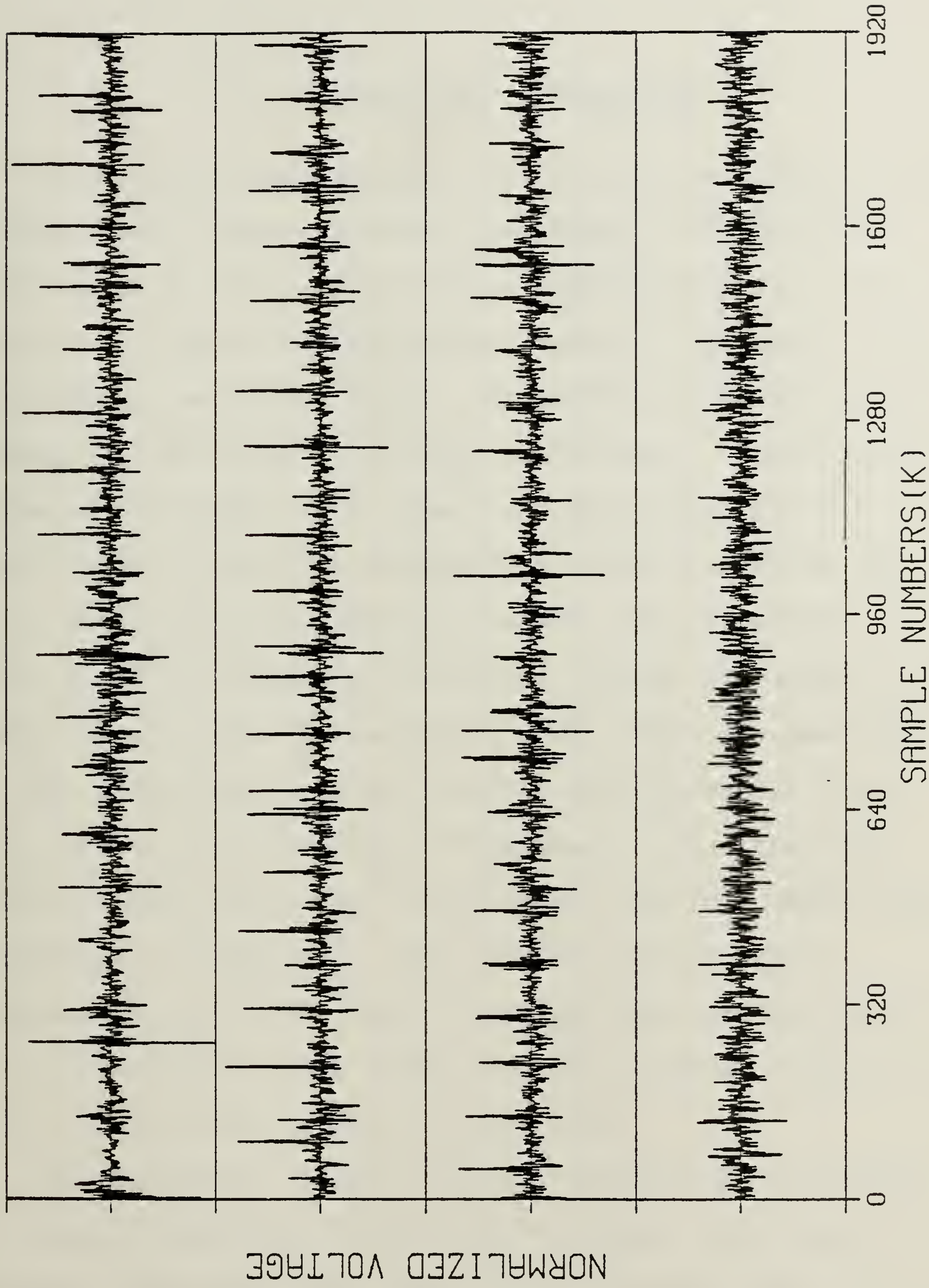

Figure 4.30. Cepstral Representation of Nine

V. RESULTS AND CONCLUSIONS

Ten speakers were selected to form the data base for the system. Their utterances were processed to obtain both their cepstral and analytic phase representations. The system was then tested using two groups of speakers. The first group, denoted Group A, consisted of speakers whose utterances were used to form the data base. Each speaker repeated the digits four times, and only three of these utterances were used to compute the average waveform and hence the cepstral and analytic phase representations. Group A can be thought of as having trained the system. The second group, Group B, consists of the other ten speakers.

The system was tested using ten utterances per digit from each of the two groups of speakers. The reference pattern space was varied, using three different spaces each containing 100 patterns. The cepstral and analytic representations formed two of the reference spaces, while the unprocessed signals formed the third space. Tables 5.1 and 5.2 contain the results of the test.

The results for Group A, in all categories, are below the results attainable with the VRM system. For three training passes the VRM system has a 97% recognition rate. The high percentage of recognition for the unprocessed

waveforms was to be expected since the speakers trained the system and the pattern space did consist of the average of each speaker's utterances. The distances between the pattern vectors and the test vector were of the same magnitude for the unprocessed waveforms, regardless of whether the utterance was correctly identified or not. In the case of the short-term phase representations when the system correctly identified an utterance, the distance between the test vector and its nearest neighbor was an order of magnitude less than all the other distances. When the system incorrectly identified an utterance all distances were of the same magnitude.

The success demonstrated in the speaker-dependent case is not without cost. As compared to the VRM system, which has at most 120 bits/pattern, this system has 122.8K bits/pattern (7680 two byte numbers). There was an extensive amount of manual editing involved to obtain these patterns, on the order of ten minutes per utterance. However, it was shown that short-term phase-only speech can be used to construct a speaker-dependent isolated word recognizer.

The results for Group B appear to be abysmal, however, several things must be considered. First, there was no preprocessing of the signals to time-wrap them. Second, no features were extracted, only the entire waveforms were

used. Third, the decision algorithm may have to be tailored to fit the data, rather than using a general purpose decision rule. Last, but certainly not least, no system exists today that is completely speaker independent.

One final observation concerning Group B. When the decision algorithm incorrectly identified any utterance it did so with a great deal of bias. In 30% of the cases where an utterance was incorrectly identified the number 'one' was picked to be the nearest neighbor.

This thesis was not an attempt to definitively answer the question, "is phase a physical invariant of speech?". Its purpose was to show that phase should be considered when constructing a word recognition system. This was accomplished. The next step is to use the information obtained from the phase in conjunction with other word recognition systems to possibly improve these systems with the long range goal of solving the speaker-independent word recognition problem.

TABLE 5.1

GROUP A RECOGNITION RESULTS
BASED ON TEN UTTERANCES PER DIGIT

Digits	Unprocessed Waveforms	Cepstral Representation	Analytic Representation
0	9	7	6
1	10	8	7
2	10	6	4
3	9	5	3
4	10	5	5
5	10	7	5
6	8	4	3
7	9	4	3
8	10	6	7
9	10	7	6
AVG	9.5	5.9	4.9

TABLE 5.2

GROUP B RECOGNITION RESULTS BASED ON TEN UTTERANCES PER DIGIT

Digits	Unprocessed Waveforms	Cepstral Representation	Analytic Representation
0	2	0	1
1	4	3	3
2	1	1	0
3	2	0	0
4	1	0	1
5	1	0	0
6	0	0	0
7	1	0	0
8	0	1	0
9	2	1	0
AUG	1.4	.6	.5

APPENDIX A

INSTRUCTION SHEET

Thank you for participating in the Speech Processing Lab's effort to collect speech samples. This exercise will require about 10 minutes of your time to complete.

I. Biographical Data

A. Name:
B. Age:
C. Sex:
D. Place of Birth:
E. Occupation:

II. Speech Sampling

A. Repeat each word on the list four times, pausing approximately 5 sec. between utterances. (For example: the first word on the list is 'zero', therefore you would say: 'zero' (pause) 'zero' (pause) 'zero' (pause) 'zero' (pause) 'one' (pause) .....)

zero
one
two
three
four
five

six
seven
eight
nine

B. Repeat the following exercise 3 times:

Read the entire list of numbers at your natural speaking rate pausing approx. 5 secs. before repeating the list. Do not pause unnaturally between the numbers. We are looking for continuous speech such as in a conversation.

zero-one-two-three-four-five-six-siven-eight-nine (pause/repeat)

APPENDIX B

COMPUTER PROGRAMS

All programs were written in IBM FORTRAN H to run on the W. R. Church Computer Center's IBM 3033. The programs access routines from the ISML library. The graphics programs interact with the DISSPLA graphics package.

```
C
      SUBROUTINE PATIN(FNI,FMI,Y)
C        (C) BY LT.JEFF PFEIFFER,JUNE 1983
C-----------------------------------------------------------------------
C          ALGORITHM READS IN A PATTERN FOR USE WITH THE DECISION
C          ALGORTHIM
C
      IMPLICIT REAL*8(A-G),INTEGER(H-M),REAL(N-Z)
      REAL Y(7680)
      INTEGER*2 IDATA(32)
C-----------------------------------------------------------------------
C .....        START THE PROGRAM
C-----------------------------------------------------------------------
      CALL FRTCMS ('FI      ','83      ','DISK    ',FNI,'CAT     ',FMI)
      REWIND 83
      DO 2010 I=1,7680,32
         READ(83,END=2030)IDATA
         DO 2020 J=1,32
            K=I+J-1
            IF (K .GT. 7680) GO TO 2020
            Y(K) = IDATA(J)
 2020    CONTINUE
 2010 CONTINUE
 2030 CONTINUE
      RETURN
      END
```

```
C                   ALGORITHM ANALYTIC
C              (C) BY LT JEFF PFEIFFER JULY 1983
C
C        COMPUTES THE ANALYTIC SIGNAL REPRESENTATION OF A SPEECH SIGNAL
C        VIA THE USE OF A 79 WEIGHT HILBERT TRANSFORM
C
         IMPLICIT INTEGER (A-Z)
         INTEGER LA,LB,IWK(10)
         REAL A(1024),B(512),Y(7680),RES(7680),H(79),ANAL(7680)
         REAL AMAX,ATEMP
         COMPLEX CA(512),CB(256),RESLT(7680)
         EQUIVALENCE (A(1),CA(1)),(B(1),CB(1))
         INTEGER*2 ANDATA(32)
         REAL*8 FN,FT,FM
C-----------------------------------------------------------------------
C .....      INITIALIZE THE DATA
C-----------------------------------------------------------------------
             LA=256
             LB=79
C
C-----------------------------------------------------------------------
C .....      READ IN THE WEIGHTS OF THE FILTER
C-----------------------------------------------------------------------
      WRITE(6,900)
 900  FORMAT('OREADING IN THE FILTER WTS,ASSUMING FN=FIR7905 FT=DAT')
      WRITE(6,910)
 910  FORMAT('OENTER THE FILEMODE')
      READ(5,920,END=930)FM
 920  FORMAT(A8)
 930  REWIND 5
      CALL FRTCMS ('FI        ','85      ','DISK    ','FIR7905 ','DAT
     & ',FM)
      REWIND 85
      DO 950 I=1,79
         READ(85,960)H(I)
 960     FORMAT(T10,F10.7)
 950  CONTINUE
C-----------------------------------------------------------------------
C .....       READ IN THE VECTOR
C-----------------------------------------------------------------------
 1    CALL INPART(1,Y)
C-----------------------------------------------------------------------
C .....     SEGMENT THE VECTOR INTO 25 MSEC SEGMENTS
C-----------------------------------------------------------------------
      DO 2000 K=1,30
C.....  ZERO THE VECTORS A & B
         DO 10 I=1,1024
            A(I) = 0.
```

```
 10       CONTINUE
          DO 20 I=1,512
             B(I) = 0.
 20       CONTINUE
C
C.....    INITIALIZE THE VECTORS A & B
C
          DO 100 I=1,256
             J=I+(K-1)*256
             A(I) = Y(J)
 100      CONTINUE
          DO 200 I=1,79
             B(I)=F(I)
 200      CONTINUE
C
C.....     CALL IMSL ROUTINE 'VCONVO' TO CONVOLVE THE VECTORS
C
          CALL VCONVO(CA,CB,LA,LB,IWK)
C
C.....     OBTAIN THE OUTPUT NOTING THAT ONLY 256 PTS OF VECTOR WERE USED
C
          DO 300 I=1,256
             J=I+(K-1)*256
             RES(J) = A(I)
 300      CONTINUE
 2000 CONTINUE
C-----------------------------------------------------------------------
C  .....   FORM THE ANALYTIC SIGNAL AND THE PHASE REPRESENTATION
C-----------------------------------------------------------------------
      DO 400 I=1,7680
          RESLT(I) = CABS(CMPLX(Y(I),RES(I)))
          ANAL(I) =(Y(I)/RESLT(I))
 400  CONTINUE
      AMAX=0.
      DO 410 I=1,7680
          ATEMP= ABS(ANAL(I))
          IF (ATEMP .GT. AMAX) AMAX=ATEMP
 410  CONTINUE
C-----------------------------------------------------------------------
C  .....        CONVERT TO I*2 FORMAT
C-----------------------------------------------------------------------
      DO 450 I=1,7680
          ANDATA(I)=(ANAL(I)/AMAX)* 10000.
 450  CONTINUE
C-----------------------------------------------------------------------
C  .....        OUTPUT THE DATA
C-----------------------------------------------------------------------
      CALL OPART(ANDATA)
```

```
C-----------------------------------------------------------------------
C .....       QUERY THE OPERATOR IF ANOTHER RUN IS DESIRED
C-----------------------------------------------------------------------
      WRITE(6,500)
 500  FORMAT('OANOTHER RUN? Y(1)N(2)')
      READ(5,550)ANS
 550  FORMAT(I1)
      GO TO (1,8000),ANS
 8000 CONTINUE
      STOP
      END
```

```
C                   ALGORITHM DECISION
C                      (C) BY LT. JEFF PFEIFFER,JULY 1983
C
C         PROGRAM COMPUTES THE EUCLIDEAN DISTANCE BETWEEN A TEST
C         VECTOR AND THE VECTORS IN THE PATTERN SPACE
C
          IMPLICIT INTEGER(I-N),REAL(C-Z),REAL*8 (F-G)
          REAL VECIN(7680),S(7680),PAT(7680)
C
C-----------------------------------------------------------------------
C .....   READ IN THE VECTOR TO BE CLASSIFIED
C-----------------------------------------------------------------------
 1      WRITE(6,100)
 100    FORMAT('0ENTER THE DIGIT U WISH TO CLASSIFY')
        READ(5,110,END=120)N
 110    FORMAT(I2)
 120    REWIND 5
        CALL INPART(1,VECIN)
C-----------------------------------------------------------------------
C       READ IN PATTERN VECTORS FROM ARRAY FORMED BY USING LISTFILE CMD
C       ASSUMING THAT THIS ARRAY HAS FILENAME = PAT,FILETYPE = DAT.
C-----------------------------------------------------------------------
        WRITE(6,130)
 130    FORMAT('0ENTER FILEMODE OF PATTERN FILE')
        READ(5,140,END=150)FM
 140    FORMAT(A8)
 150    REWIND 5
        CALL FRTCMS('FI      ','82      ','DISK    ','PAT     ','DAT
     a',FM)
        REWIND 82
        DO 200 I=1,100
          READ(82,210)FNI,FMI
 210      FORMAT(T6,A8,T26,A8)
          CALL PATIN(FNI,FMI,PAT)
C-----------------------------------------------------------------------
C .....             COMPUTE THE EUCLIDEAN DISTANCE
C-----------------------------------------------------------------------
            SUM = 0.
            DO 260 J=1, 7680
               SUM=SUM+ ( (PAT(J)-VECIN(J))/10000.)**2
 260        CONTINUE
            S(I) = SQRT (SUM)
 200      CONTINUE
C-----------------------------------------------------------------------
C .....          OUTPUT THE RESULTANT VECTOR
C-----------------------------------------------------------------------
        WRITE(6,300)
 300    FORMAT('0FILENAME OF RESULTANT VECTOR')
```

```
      READ(5,310,END=320)FNR
 310  FORMAT(A8)
 320  REWIND 5
      WRITE(6,330)
 330  FORMAT('0FILEMODE OF RESULTANT VECTOR')
      READ(5,310,END=340)FMR
 340  REWIND 5
      CALL FRTCMS('FI      ','80      ','DISK    ',FNR,'DEC     ',FMR)
      REWIND 80
      WRITE(80,490)
 490  FORMAT('1')
      WRITE(80,500)N
 500  FORMAT(T15,'NUMBER TO BE CLASSIFIED IS ',I2)
      DO 510 I=1,100
         WRITE(80,520)I,S(I)
 520     FORMAT(T5,I2,T15,E14.7)
 510  CONTINUE
C-----------------------------------------------------------------------
C .....    QUERY OPERATOR FOR ANOTHER RUN
C-----------------------------------------------------------------------
      WRITE(6,540)
 540  FORMAT('0ANOTHER RUN? Y(1),N(2)')
      READ(5,550)NANS
 550  FORMAT(I1)
      GO TO (1,2000),NANS
2000  CONTINUE
      STOP
      END
```

```
C                        ALGORITHM FREQUENCY RESPONSE
C                          (C) MAY 1983 J.T.PFEIFFER
C
C          THIS IS A DOUBLE PRECISION ALGORITHM TO COMPUTE THE
C          FREQ RESPONSE OF A F.I.R. HILBERT TRANSFORM WITH ODD SYMMETRY
C          ABOUT THE POINT (N-1)/2.  THE RESPONSE IS CALCULATED FOR THE
C          FREQUENCIES FROM 0 TO PI RADIANS.  THE OUTPUT PLOT USES
C          THE DISSPLA ROUTINE AND USES AS THE ABSCISSA THE
C          NORMALIZED FREQUENCY (I.E. FRACTION OF 2 PI)
C
C         * * * * * * * * *    FILEDEF'S USED   * * * * * * *
C             10 - TERMINAL
C             05 - INPUT WEIGHTS
C
C         * * * * * * VARIABLES DEFINED * * * * * * * * * * *
C          N - NUMBER OF WEIGHTS,INTEGER
C          H - VECTOR OF WEIGHTS, REAL
C          M - (N-1)/2,INTEGER
C          HSTAR - VALUE OF THE REAL PART OF THE RESPONSE,REAL
C          E - REAL PART OF RESPONSE,REAL
C          F - IMAGINARY PART OF RESPONSE,REAL
C          C - COMPLEX VECTOR (E,F).COMPLEX
C          X - ABSCISSA FOR PLOTTING,REAL
C          Y - ORDINATE FOR PLOTTING,REAL
C          D - INCREMENT IN FREQUENCY, REAL
C          DEL - PEAK APPROX ERROR,REAL
C          FL - LOWER CUTOFF FREQ,REAL
C
C         * * * * * * * DECLARE VARIABLES * * * * * * * * *
C
      IMPLICIT COMPLEX(A-C),INTEGER(I-P),REAL(Q-Z)
      REAL*8 PI,HSTAR,ERROR,D,H
      DIMENSION H(900),X(900),Y(900),C(900),ERROR(900)
      INTEGER SWITCH,III,ZZ,CC
      REAL XDIM,YDIM,XPG,YPG,GR,HT,F,E,II,XX(900),FH,FL,DELM,KK,DEL
C
C         * * * * * * * * INITIALIZE VARIABLES   * * * * * * *
C
      PI = (3.14159265 )
C
C         * * * * * * * * BEGIN ALGORITHM * * * * * * * * * *
C
C         * * * READ IN SPECIFICATIONS FOR THE FILTER * * * *
C
      WRITE(10,100)
  100 FORMAT('ENTER NUMBER OF WEIGHTS TO BE USED, IN I2 FORMAT')
      READ(5,110)N
  110 FORMAT(I2)
```

```
      WRITE(10,112)
112   FORMAT('ENTER THE VALUE OF THE PEAK APPROX ERROR,D,IN F4.2 FORMAT'
     a)
      READ(5,114)DEL
114   FORMAT(F4.2)
      WRITE(10,115)
115   FORMAT('ENTER THE LOWER CUTOFF FREQ FL  IN F4.2 FORMAT')
      READ(5,116)FL
116   FORMAT(F4.2)
      WRITE(10,117)
117   FORMAT('ENTER THE NUMBER OF PAIRS OF POINTS TO PLOT,IN I3 FORMAT')
      READ(5,118)P
118   FORMAT(I3)
C
C        * * * * * READ IN VALUES OF WEIGHTS * * * * * * * *
C
      DO 10  I=1,N
         READ(85,120)H(I)
120      FORMAT(I10,F10.7)
10    CONTINUE
C
C        * * * * * COMPUTE THE VALUE OF THE IMPLUSE RESPONSE * *
C
      G = FLOAT(F)
      L = P+1
      M=(N-1)/2
      DO 20  I=1,L
         HSTAR=0.
         D = FLOAT(I-1)*(PI/G)
         X(I) = FLOAT(I-1)/(2.*G)
         DO 30 J=1,M
            HSTAR = HSTAR + 2.*H(M-J+1)*DSIN(D*J)
30       CONTINUE
         E = HSTAR*DSIN(M*D)
         F = HSTAR*DCOS(M*D)
         C(I)= CMPLX(E,F)
20    CONTINUE
C
C        * * * COMPUTE MAGNITUDE OF THE FREQ RESPONSE  * * *
C
      DO 40  I= 1,L
         Y(I)= CABS(C(I))
40    CONTINUE
C
C        * * * COMPUTE THE ERROR OVER THE RANGE (FL,FH)* * * *
C
      DELM = -DEL
      II = 2.*G*FL + 1.
```

```
      III = INT(II)
      FH = .5 - FL
      KK = 2.*G*FH + 1.
      K = INT(KK)
      DO 50 I = III,K
         ZZ = I + 1 -III
         ERROR(ZZ) = (Y(I) - 1.)
         XX(ZZ) = X(I)
 50   CONTINUE
      CC = K - III + 1
C
C       * * * * * * * * * PLOT OUTPUT * * * * * * * * * * *
C
C       * * * * PARAMETERS FOR DISSPLA  * * * * * * * * * *
C
      WRITE(10,300)
 300  FORMAT('ENTER 1 IF TEK 618 IS USED,ENTER 2 IF OTHER DEVICE')
      READ(5,310)SWITCH
 310  FORMAT(I1)
      WRITE(10,320)
 320  FORMAT('ENTER X-DIM FOR PAGE,F4.1')
      READ(5,330)XDIM
 330  FORMAT(F4.1)
      WRITE(10,340)
 340  FORMAT('ENTER Y-DIM FOR PAGE,F4.1')
      READ(5,350)YDIM
 350  FORMAT(F4.1)
      WRITE(10,360)
 360  FORMAT('ENTER X-DIM FOR SUBPLOT AREA,F4.1')
      READ(5,370)XPG
 370  FORMAT(F4.1)
      WRITE(10,380)
 380  FORMAT('ENTER Y-DIM FOR SUBPLOT AREA,F4.1')
      READ(5,390)YPG
 390  FORMAT(F4.1)
      WRITE(10,400)
 400  FORMAT('ENTER GRACE MARGIN,F3.1')
      READ(5,410)GR
 410  FORMAT(F3.1)
      WRITE(10,420)
 420  FORMAT('ENTER CHAR HEIGHT SPEC,F3.1')
      READ(5,430)HT
 430  FORMAT(F3.1)
C
C
C        * * * * * SET UP LARGE BUFFER * * * * * * * * * *
      CALL LRGBUF
C
```

```
C        * * * * * * * CALL DEVICE TO BE USED * * * * * * * *
C
      GO TO (500,600),SWITCH
 500  CALL TEK618
      GO TO 620
 600  CALL PTEKAL
 620  CONTINUE
C
C        * * * * * * * * SET UP PAGE* * * * * * * * * * *   * *
C
      CALL HEIGHT(HT)
      CALL PAGE(XDIM,YDIM)
      CALL NOBRDR
      CALL GRACE (GR)
C
C         * * * * * * CALL SETUP FOR SUBPLOT AREA * * * * * * *
C
      CALL AREA2D(XPG,YPG)
C
C         * * * * * * * NAME THE X & Y AXISES * * * * * * * *
C
      CALL XNAME('FREQUENCY NORMALIZED TO 2 PI$',100)
      CALL YNAME('MAGNITUDE$',100)
      CALL YAXANG(0.)
C
C        * * * * * * SET UP HEADING FOR THE GRAPH * * * * *  *
C
      CALL HEADIN('FREQ RESPONSE OF HILBERT TRANSFORM WITH 79 WEIGHTS$',
     a100,1.,3)
      CALL HEADIN('PEAK APPROX ERROR = 0.0388830$',100,1.,3)
      CALL HEADIN('LOWER CUTOFF FREQ = 0.01$',100,1.,3)
C
C        * * * * * * * SET UP THE GRAPH SPECS * * * * * * * *
C
      CALL GRAF(0.,.25,.5,0.,.05,1.05)
C
C        *  * * * * * * * DRAW THE GRAPH * * * * * * * * * * *
      CALL CURVE(X,Y,0,0)
C
      CALL ENDPL(0)
C
C          * * * * * * * START FOR THE SECOND GRAPH * * * * * * *
C
      WRITE(10,720)
 720  FORMAT('ENTER X-DIM FOR PAGE,F4.1')
      READ(5,730)XDIM
 730  FORMAT(F4.1)
      WRITE(10,740)
```

```
740   FORMAT('ENTER Y-DIM FOR PAGE,F4.1')
      READ(5,750)YDIM
750   FORMAT(F4.1)
      WRITE(10,760)
760   FORMAT('ENTER X-DIM FOR SUBPLOT AREA,F4.1')
      READ(5,770)XPG
770   FORMAT(F4.1)
      WRITE(10,780)
780   FORMAT('ENTER Y-DIM FOR SUBPLOT AREA,F4.1')
      READ(5,790)YPG
790   FORMAT(F4.1)
      WRITE(10,800)
800   FORMAT('ENTER GRACE MARGIN,F3.1')
      READ(5,810)GR
810   FORMAT(F3.1)
      WRITE(10,820)
820   FORMAT('ENTER CHAR HEIGHT SPEC,F3.1')
      READ(5,830)HT
830   FORMAT(F3.1)
C
C        * * * * * * * * SET UP PAGE * * * * * * * * * * *
      CALL HEIGHT(HT)
      CALL PAGE(XDIM,YDIM)
      CALL NOBRDR
      CALL GRACE (GR)
C
C     * * * * * * CALL SETUP FOR SUBPLOT AREA * * * * * * * *
      CALL AREA2D(XPG,YPG)
C
C     * * * * * * * NAME THE X & Y AXISES * * * * * * * * * * * *
C
      CALL XNAME('FREQUENCY NORMALIZED TO 2 PI$',100)
      CALL YNAME('MAGNITUDE$',100)
      CALL YAXANG(0.)
C
C     * * * * * *  SET UP HEADING FOR THE GRAPH * * * * * * * *
C
      CALL HEADIN('ERROR OF HILBERT TRANSFORM WITH 79 WEIGHTS$',100,1.,3
     a)
      CALL HEADIN('PEAK APPROX ERROR = 0.0388830$',100,1.,3)
      CALL HEADIN('LOWER CUTOFF FREQ = 0.01$',100,1.,3)
C
C     * * * * * * * * * SET UP THE GRAPH SPECS * * * * * * * * * *
C
      CALL GRAF(FL,'SCALE',FH,DELM,'SCALE',DEL)
C
C     * * * * * * * *  DRAW THE GRAPH * * * * * * * * * *
C
```

```
      CALL CURVE(XX,ERRCR,CC,0)
C
C        * * * * * *  ENC THE PLOT* * * * * * * * * * * *
C
      CALL ENDPL(C)
      CALL DCNEPL
      STCP
      ENC
```

```
C                 ALGORITHM GRAPH
C          THIS PROGRAM IS A MODIFICATION OF A GRAPHICS ROUTINE,GRAF1,
C          WRITTEN BY LT JAY H. BENSON IN MAY 1983.  MODIFICATION WAS
C          DONE BY LT JEFFREY T. PFEIFFER IN JULY 1983.
C
C            PROGRAM PLOTS 7680 SAMPLES OF SPEECH DATA
C-----------------------------------------------------------------------
      IMPLICIT INTEGER (A-V),REAL(X-Z)
      INTEGER*2 IDATA(32)
      LOGICAL*1 HEAD1(25)
      LOGICAL*1 LTRDOL, LTRBLK
      REAL X(7680), Y(7680)
      REAL Y1(1920), Y2(1920), Y3(1920), Y4(1920), Y5(1920)
      DATA LTRDOL/'$'/, LTRBLK/' '/
      EQUIVALENCE (Y(1),Y1(1))
      EQUIVALENCE (Y(1921),Y2(1))
      EQUIVALENCE (Y(3842),Y3(1))
      EQUIVALENCE (Y(5761),Y4(1))
C-----------------------------------------------------------------------
C.....    CALL THE EXEC GR TO ESTABLISH THE FILEDEFS
C-----------------------------------------------------------------------
1000  CONTINUE
      CALL FRTCMS ('EXEC ','GR    ')
      READ (5,901,END=8000) IANS
901   FORMAT (I1)
      IF (IANS.NE.0) GOTO 8000
C-----------------------------------------------------------------------
C.....    READ IN THE WAVEFORM TO BE PLOTTED
C-----------------------------------------------------------------------
      REWIND 1
      DO 1010 I = 1,7680,32
      READ (1,END=1030) IDATA
      DO 1020 J = 1,32
      K = I + J -1
      IF (K.GT.7680) GOTO 1020
      Y(K) = IDATA(J)
1020  CONTINUE
1010  CONTINUE
1030  CONTINUE
C-----------------------------------------------------------------------
C.....     NORMALIZE THE DATA TO BE PLOTTED
C-----------------------------------------------------------------------
      IMAX = 0
      DO 2000 I = 1,7680
      TEMP = ABS(Y(I))
2000  IF (TEMP.GT.IMAX) IMAX = TEMP
      DO 2010 I = 1,7680
      X(I) = I
```

```
2010  Y(I) = Y(I) / IMAX
C-----------------------------------------------------------------------
C.....   QUERY THE OPERATOR FOR GRAPHICS PARAMETERS
C-----------------------------------------------------------------------
      WRITE(6,160)
 160  FORMAT('OPTEKAL(1),TEK618(2)?')
      REWIND 5
      READ(5,170,END=400)DEV
 170  FORMAT(I1)
 400  CONTINUE
      WRITE(6,180)
 180  FORMAT('OCOMPRS(1),NO(2)?')
      REWIND 5
      READ(5,190,END=405)CNUM
 190  FORMAT(I1)
 405  CONTINUE
      WRITE(6,200)
 200  FORMAT('OENTER BLOWUP FACTOR,F3.1')
      REWIND 5
      READ(5,210,END=410)XFACT
 210  FORMAT(F3.1)
 410  CONTINUE
      IF (DEV .EQ. 1) CALL PTEKAL
      IF (DEV .EQ. 2) CALL TEK618
      IF (CNUM .EQ. 1) CALL COMPRS
C-----------------------------------------------------------------------
C..... QUERY OPERATOR AS TO THIRD LINE OF HEADING
C-----------------------------------------------------------------------
      DO 2100 I = 1,24
      HEAD1(I) = LTRBLK
2100  CONTINUE
      HEAD1(25) = LTRDOL
      WRITE (6,910)
910   FORMAT ('OENTER SUBHEADING LINE 3 OF 3 (END WITH $ TO CENTER):')
      REWIND 5
      READ (5,911,END=2110) (HEAD1(I),I=1,24)
911   FORMAT (24A1)
2110  CONTINUE
C-----------------------------------------------------------------------
C.....   SETUP THE PLOTTING AREA
C-----------------------------------------------------------------------
      XPAGE = 11.0
      YPAGE = 8.5
      CALL PAGE (XPAGE,YPAGE)
      CALL NOBRDR
      XAXIS = 8.0
      YAXIS = 5.0
      CALL AREA2D (XAXIS, YAXIS)
```

```
      CALL BLOWUP(XFACT)
C-----------------------------------------------------------------------
C.....    LABEL THE X & Y AXISES
C-----------------------------------------------------------------------
      CALL XNAME ('SAMPLE NUMBERS(K)$',100)
      CALL YNAME ('NORMALIZED VOLTAGE$',100)
      CALL HEADIN ('SAMPLED SPEECH WAVEFORM$',100,1.3,3)
      CALL HEADIN ('UTTERANCE IS$',100,1.1,3)
      CALL HEADIN (HEAD1,100,1.1,3)
C-----------------------------------------------------------------------
C.... .    DEFINE THE AXISES
C-----------------------------------------------------------------------
      XORIG = 0.0
      XSTP = 320.0
      XMAX = 1920.0
      YORIG =-1.0
      YSTP = 1.0
      YMAX = 7.0
      CALL YNONUM
      CALL GRAF (XORIG,XSTP,XMAX,YORIG,YSTP,YMAX)
      CALL GRID (0,1)
C-----------------------------------------------------------------------
C.....   SEGMENT THE WAVE INTO 4 PARTS OF 1920 PTS. EACH.  GRAPH
C        ALL THE SEGMENTS ON ONE PHYSICAL PAGE BEGINNING WITH
C        SEGMENT 1 BIASING THE SEGMENTS AS FOLLOWS:
C           SEGMENT 1 BY+6.0
C           SEGMENT 2 BY+4.0
C           SEGMENT 3 BY+2.0
C           SEGMENT 4 BY+0.
C-----------------------------------------------------------------------
      DO 3000 I = 1,1920
3000  Y(I) = Y(I) + 6.0
      DO 3010 I = 1921,3840
3010  Y(I) = Y(I) + 4.0
      DO 3020 I = 3841,5760
3020  Y(I) = Y(I) + 2.0
C-----------------------------------------------------------------------
C...  DRAW THE FIVE CURVE SEGMENTS
C-----------------------------------------------------------------------
      NPNTS = 1920
      IMARK = 0
      CALL CURVE (X,Y1,NPNTS,IMARK)
      CALL CURVE (X,Y2,NPNTS,IMARK)
      CALL CURVE (X,Y3,NPNTS,IMARK)
      CALL CURVE (X,Y4,NPNTS,IMARK)
      CALL CURVE (X,Y5,NPNTS,IMARK)
C-----------------------------------------------------------------------
C.....    TERMINATE THIS PLOT
```

```
C-----------------------------------------------------------------------
      CALL ENDPL (0)
C-----------------------------------------------------------------------
C.....     LCOP EACK FOR ANCTHER PLOT, IF DESIRED
C-----------------------------------------------------------------------
      GCTO 1COO
C-----------------------------------------------------------------------
C.....     NCRMAL EXIT
C-----------------------------------------------------------------------
ECCC  CCNTINLE
      CALL DCNEPL
      RETURN
      ENC
```

LIST OF REFERENCES

1. Flanagan, J. L., "Voices of Men and Machines," Speech Analysis, Schafer, Ronald W. and Markel, John D., ed., p. 4-16, IEEE Press, 1979.

2. Flanagan, J. L., Speech Analysis Synthesis and Perception, p. 1-35, Springer-Verlag, 1972.

3. Reddy, D. Raj, "Speech Recognition by Machine," Automatic Speech and Speaker Recognition, Dixon, N. Rex and Martin, Thomas B., ed., p. 56-86, IEEE Press, 1979.

4. Interstate Electronics Corporation, Voice Recognition Module Reference Manual, 1980.

5. Rome Air Development Center, RADC/NATO Word Recognition Tests, p. 1-25, October 1980.

6. Ibid., p. 12.

7. Schroeder, Manfred R., "Models of Hearing," Proceedings of the IEEE, v. 63, p. 1332, September 1975.

8. Hudspeth, A. J., "The Hair Cells of the Inner Ear," Scientific American, p. 54-64, January 1983.

9. Schroeder, Manfred R., p. 1344.

10. Ibid., p. 1345.

11. Ibid., p. 1346.

12. Ibid., p. 1347.

13. Ibid., p. 1348.

14. Oppenheim, Allan V. and others, "Phase in Speech and Pictures," Proceedings of the 1979 IEEE Conference on Acoustics, Speech and Signal Processing, p. 632-635, April 1979.

15. Cox, R. C. and Robinson, D. M., "Some Notes on Phase in Speech Signals," Proceedings of the 1980 IEEE Conference on Acoustics, Speech, and Signal Processing, p. 150-153, April 1980.

16. Rabiner, Lawrence R. and Gold, Benjamin, *Theory and Application of Digital Signal Processing*, p. 650-690, Prentice-Hall, 1970.

17. Ibid., p. 663.

18. Ibid., p. 688.

19. Ibid., p. 689.

20. Shankar, R. and McDonough, R. N., "Ultrasonic Measurements of Defects in Metals Using Cepstral Processing," *Proceedings of the 1978 IEEE Conference on Acoustics, Speech, and Signal Processing*, p. 533-537, April 1978.

21. Cox, R. C. and Robinson, D. M., p. 151.

22. Ibid., p. 152.

23. Rabiner, L. R. and Schafer, R. W., "On the Behavior of Minimax FIR Digital Hilbert Transformers," *Bell System Technical Journal*, v. 53, p. 363-389, February 1974.

24. Ibid., p. 372.

25. Cox, R. C. and Robinson, D. M., p. 152.

26. Rabiner, L. R. and Schafer, R. W., p. 376-379.

27. GENRAD, *2500-Series System*, 1981.

28. Rome Air Development Center, *Keyword Operational Analysis*, by S. Moshier and L. Bahler, p. 55, August 1981.

29. Therrien, C. W., "Notes for an Introductory Course in Pattern Recognition," Notes at the Naval Postgraduate School, Monterey, California, 1982.

BIBLIOGRAPHY

Caprania, Robert R., The Evoked Vocal Response of the Bullfrog: A Study of Communication by Sound, M.I.T. Press, 1965.

Chatfield, C., The Analysis of Time Series, Chapman and Hall, 1972.

Oppenheim, Alan V., Applications of Digital Sound Processing, Prentice-Hall, 1978.

Roberts, Russ, Signal Processing Techniques, Interstate Electronics Corporation, 1977.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943	2
3.	Department Chairman, Code 62 Department of Electrical Engineering Naval Postgraduate School Monterey, California 93943	1
4.	Professor Steven Jauregui, Code 62Ja Department of Electrical Engineering Naval Postgraduate School Monterey, California 93943	10
5.	Associate Professor Alex Gerba, Code 62Gz Department of Electrical Engineering Naval Postgraduate School Monterey, California 93943	1
6.	Professor S. R. Parker, Code 62Px Department of Electrical Engineering Naval Postgraduate School Monterey, California 93943	1
7.	Commander Naval Security Group Command Naval Security Group Command Headquarters 3801 Nebraska Avenue, N.W. (ATTN: Code G82) Washington, D.C. 20390	2
8.	Commander Naval Electronics Systems Command Naval Electronics Systems Command Headquarters (ATTN: Code PME-107-9) Washington, D.C. 20360	2

9. Director, National Security Agency 1
Group R
Fort George G. Meade, Maryland 20755

10. Dr. Bart Rice, Code R6 1
National Security Agency
Fort George G. Meade, Maryland 29755

11. Director, National Security Agency 1
Group W3
Fort George G. Meade, Maryland 20755

12. LT J. H. Benson 1
Department of Defense, Code R211
9800 Savage Road
Fort George G. Meade, Maryland 20755

13. LT S. J. Levanduski 1
1220 Spruance Road
Monterey, California 93940

14. LT R. Lyman 1
1061 Halsey Drive
Monterey, California 93940

15. CAPT Charles Shawcross 1
2379 Irving Avenue
Monterey, California 93940

16. LT J. T. Pfeiffer 1
942 Autumnwood Drive
Gambrills, Maryland 21054

207539

Thesis
P46164 Pfeiffer
c.1 The importance of
phase in word recognition.